AN INVESTIGATION INTO THE RNA-BINDING PROTEIN UNR AND ITS INTERACTORS

by

Pól Roibeárd Ó CATNAIGH

Thesis submitted in partial fulfilment of the requirements for the degree of:

Doctor of Philosophy in Systems Biology

University of Warwick,

Department of Computer Science,

December 2016

LIST OF TABLES:

LIST OF FIGURES:

Acknowledgements

I wish to state my immense gratitude to my PhD supervisors, Drs Emma Anderson and Sascha Ott, without whose ongoing help and encouragement this thesis would never have been completed. I could not have hoped for more patient, supportive, knowledgeable and affable supervisors. Whilst Emma provided me with more day to day support over my time in her lab, Sascha always made himself available to help with his theoretical expertise. He was particularly helpful in checking my draft thesis at the end of my course when Emma was unfortunately on long-term sick leave.

I would also like to thank the staff and students here in the virology department of the School of Life Sciences who helped me in various ways throughout my course, in particular Emma's former postdoc, Dr Swagat Ray. I am likewise grateful to the staff and students in the, now defunct, Systems Biology Doctoral Training Centre, in particular to Sascha's former postdoc, Dr Nigel Dyer. Finally, I wish to extend my gratitude to various members of support staff who helped by running my mass spectrometry samples and part-processing my RNA-Seq samples, as well as to Cerith Harries in the media preparation department.

Outside of Warwick, I would like to thank my amazing wife for being there to help and encourage me throughout the past few years. Hopefully we can now start to enjoy the life together that was put on hold during the production of this document!

Declaration

This work was completed at the University of Warwick between October 2012 and December 2016, and has not been submitted for another degree. I declare that the work is original and, unless otherwise stated in the text, has been completed by the author.

Pól Roibeárd Ó CATNAIGH December 2016

Abstract

Recent work linked the RNA-binding protein UNR to a number of human pathologies although little is currently known about how UNR functions in human cells. This thesis aims to elucidate how UNR functions by, among other things, discovering novel UNR-interacting proteins and transcripts and proteins that are differentially expressed in the presence or absence of UNR.

It is shown that UNR levels decrease with increasing cell confluency in cultured HeLa cells but that they increase with increasing confluency in the wild type TP53-containing U2OS cell line. UNR is shown to co-localise to stress granules with TP53 in arsenite-stressed HeLa cells.

A number of novel UNR-interacting proteins were discovered in three cell lines (HeLa, U2OS and SaOS-2), including HUWE1, NARR, SQSTM1 and LDB1. GO-term overrepresentation analysis confirmed that UNR is an RNA-binding protein as 'RNA binding' and 'poly(A) RNA binding' were the top two overrepresented molecular function GO terms by p-value across each of the cell types. Less expected overrepresented GO terms pertained to selenium metabolism and the extracellular exosome.

There was no evidence for conservation of UNR-interacting transcripts across the cell types but there were some similar significantly overrepresented GO terms among the respective UNR-interacting transcripts. These included terms pertaining to RNA and the nucleus. The most significant UNR-interacting transcript in HeLa cells was *PABPC1* and that the PABP protein was also significantly upregulated following *UNR* knockdown in HeLa cells.

'Poly (A) RNA binding' was a significantly overrepresented GO term among proteins differentially regulated following *UNR* knockdown in HeLa and U2OS cells. 'Adherens junction' was another significantly overrepresented GO term using proteins that were higher in abundance in si*UNR*-treated HeLa cells and either higher or lower in abundance in either arsenite-stressed or unstressed U2OS cells. UNR was observed at cell-cell junctions in HeLa and U2OS cells by immunofluorescence microscopy.

List of Abbreviations

ARE	AU-rich element
BLAT	Basic Local Alignment Search Tool [BLAST]-Like Alignment Tool
BSA	Bovine Serum Albumin
CITE	Cap-Independent Translational Enhancer
CONON	Copy Number Analysis
COSMIC	Catalogue Of Somatic Mutations In Cancer
CSD	Cold Shock Domain
CSDE1	Cold Shock Domain-containing protein E1
DBA	Diamond Blackfan Anaemia
DMEM	Dulbecco's Modified Eagle's Medium
DTT	Dithiothreitol
dUnr	*Drosophila melanogaster* UNR
ECL	Enhanced ChemiLuminescence
EDTA	2,2',2'',2'''-(Ethane-1,2-diyldinitrilo)tetraacetic acid
EMT	Epithelial-Mesenchymal Transition
FCS	Foetal Calf Serum
GO	Gene Ontology
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HRP	Horseradish Peroxidase
HUWE1	HECT, UBA and WWE domain containing protein 1
IF	ImmunoFluorescence (microscopy)
IP	ImmunoPrecipitation
IRES	Internal Ribosome Entry Site
ITAF	Internal Ribosome Entry Site (IRES) Trans-Acting Factor
LDB1	LIM Domain Binding protein 1
mCRD	Major protein-Coding-Region Determinant of instability
MHC	Major Histocompatibility Complex
MOMP	Mitochondrial Outer Membrane Permeablisation
NARR	Nine-Amino acid Residue-Repeats
NMJ	NeuroMuscular Junction
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PLB	Polysome Lysis Buffer
ψCSD	PSeudo Cold Shock Domain
RIP	Ribonucleoprotein ImmunoPrecipitation
RNP	RiboNucleoProtein (domain)
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SQSTM1	Sequestosome-1
TGCC	Testicular Germ Cell Cancer
UNR	Upstream of N-Ras
UNRIP	Upstream of N-ras (UNR)-Interacting Protein
UTR	Untranslated Region
VRC	Vanadyl Ribonucleoside Complex

1 Introduction

The flow of genetic information from DNA to RNA to protein has been referred to as the Central Dogma of Molecular Biology although the author of this term later qualified this simplistic concept of unidirectional flow (Crick, 1970) following the discovery of reverse transcription in RNA viruses (Baltimore, 1970; Temin & Mizutani, 1970). Although the idea of unidirectional flow from DNA to RNA and ultimately to protein is not a universal truth, it is nevertheless true that the processes of transcription and translation can generate protein from DNA through an mRNA intermediate. RNA occupies a central role in this process and is a key point of control in protein synthesis. As well as messenger RNA (mRNA), transfer RNAs and ribosomal RNAs are fundamental to protein synthesis. Other non-coding RNAs can play a role in various processes relating to protein synthesis, including RNA degradation (considered in Cech & Bass 1986; He & Hannon 2004), repression of translation (He & Hannon, 2004), and splicing (Lerner, Boyle, Mount, Wolin, & Steitz, 1980; Waring & Davies, 1984). RNA-binding proteins have a variety of functions in the cell, including splicing (Gubitz, Feng, & Dreyfuss, 2004), RNA export from the nucleus (Erkmann & Kutay, 2004), RNA localisation within the cell (Holt & Bullock, 2009) and translation initiation (Sonenberg & Hinnebusch, 2009). Various groups of RNA-binding domains exist within RNA-binding proteins that allow them to bind to RNAs to carry out their roles. These include RNA-recognition motifs (Cléry, Blatter, & Allain, 2008), cold shock domains (Mihailovich, Militti, Gabaldón, & Gebauer, 2010) and zinc finger domains (Hall, 2005).

This thesis was undertaken in order to expand upon both the current knowledge of the cold shock domain-containing RNA-binding protein 'upstream of *NRAS*' (UNR)-interacting proteins and transcripts and upon the current knowledge of proteins that have their expression levels modulated by UNR. It is laid out in seven chapters.

This first introductory chapter considers the current state of the literature concerning *UNR*/UNR. It first considers the *UNR* gene, then its transcripts and finally the UNR protein. The second chapter lists materials and methods used in the thesis. The third chapter is an introductory experimental chapter in which the distribution of UNR within cells under different conditions is explored together with the expression levels of UNR at different cell confluencies. It is shown that UNR levels are inversely related to cell confluency in HeLa cells although the relationship is reversed in U2OS cells. The fourth chapter explores novel UNR-interacting proteins and suggests some that appear to be consistent across three different cell lines (HeLa, U2OS and SaOS-2). The

fifth chapter explores novel UNR-interacting transcripts with particular reference to groups of transcripts that are involved in a biological process, molecular function or that are located within a given cellular component. The sixth chapter explores the effect of knocking down *UNR* upon the proteome. A conclusion then follows that draws together salient findings from the earlier chapters.

1.1 The *UNR* gene

1.1.1 Nomenclature

When Northern blot analysis of guinea pig *NRAS* led Doniger & DiPaolo to suggest that another gene was located immediately upstream of *NRAS*, they conferred the locative name 'upstream of *NRAS*' (*UNR*) upon it (Doniger & DiPaolo, 1988). Indeed, the intimate proximity of these two genes is unrivalled within many genomes; their coding sequences are separated by a mere 130 bases in mice (Jeffers, Paciucci, & Pellicer, 1990). The *UNR* gene is also known by a variety of other names, such as the synteny mapping-based 'D1S155E'. Such names point to the genomic location of *UNR* but gave no information about its function. The term 'Cold Shock Domain-containing protein E1' (CSDE1) is the preferred name for the UNR protein in the www.uniprot.org database. This name refers to the structure of the protein and, by inference, suggests that it may function in a way akin to other cold shock domain containing proteins documented in the literature. This chapter will address the current literature on UNR/UNR, mainly focussing on the protein, and will allude to some exciting research being undertaken in the Anderson lab that lies beyond the scope of this thesis.

1.1.2 Conservation of *UNR*

The *UNR* gene is widely conserved among the metazoan linage (Mihailovich et al., 2010). Interestingly, the nematode worm, *Caenorhabditis elegans*, appears to have secondarily lost its *UNR* homologue over evolutionary time (Mihailovich et al., 2010). Such observations as these imply both an early origin and an essential function for *UNR*. The observed intimacy between the coding regions of *NRAS* and *UNR* in the guinea pig genome was later shown to be widely

conserved within mammalian genomes (Jacquemin-Sablon et al., 1994). For example, there are around 150 bases between the human genes (Jacquemin-Sablon & Dautry, 1992).

1.1.3 Transcriptional interference on expression of murine *Nras*

The spatial proximity between the genes may result in *UNR* exerting transcriptional interference on *NRAS* (Boussadia et al., 1997). To test this hypothesis, mice were bred that were heterozygous for deletion of the *Unr* promoter. These rodents expressed a significantly reduced amount of *Unr* transcript across a number of different tissues. As expected, the reduction in *Unr* mRNA levels were accompanied with concomitant increases in *Nras* transcript levels. Interestingly, however, the absolute amount the *UNR* transcript in given tissues, taken to be proportional to the rate of *UNR* transcription, did not appear to be related to the level of *NRAS* mRNA in those same tissues. This is considered further in section 1.5.4.

1.1.4 *Unr* is an essential gene in mice

It was noted that no *Unr* -/- homozygotes were born of heterozygote parents, leading the investigators to surmise that *Unr* is essential in mice (Boussadia et al., 1997). As discussed below, recent work has helped to explain this observation on a functional level although much remains to be elucidated in relation to the molecular mechanism by which Unr affects embryonic viability in the mouse (Elatmani et al., 2011).

1.1.5 UNR copy number

It was suggested that there is only one murine *Unr* gene and that it is located in close proximity to murine *Nras*. The human genome, however, was predicted to possess two *UNR* sequences, of which one was believed to be situated immediately upstream of *NRAS* and the other at an unknown location elsewhere in the genome (Jeffers et al., 1990). It should be pointed out that there is currently little evidence in the literature to support the existence of a second, non-*NRAS*-related, *UNR* in the human genome.

In their attempt to ascertain the murine and human *Unr*/*UNR* genetic copy numbers, Jeffers *et al*. (1990) used the following cDNA sequence as a probe:

GATCCTTGGTGCAGCTTCTGTTCAACTTTGTATCACGGGAACGGATTGGCCTGATTTCTTGGCC
TTCCTCTTGAATTGGCCCCAAACAGGGTCCCTGGCAAGTGGAGTGAAGGCTTTTTGTCTAAAGA
TGACAAGGGTCAGCTCAGGGGTTGTGGGGGAGGGCGTTTTCATCTTCCCCGTTGTCACTTGAGG
TTTTGAACTCTGGGTAAAGAGGCCGTTTATCTTTGTAAACACAAAACATTTTTGCTTTCTCCGG
TTTCATGTTAATGGCGAAAGAATGGAAGCgaataaaGTTTTACTGATTTTTGAAAAAAAAAGGA
ATTC

It equates to part of the 3'UTR of the rat *Unr* transcript. The sequence is 97.1% identical to the mouse equivalent and 90.5% identical to the human equivalent, according to a Basic Local Alignment Search Tool [BLAST]-Like Alignment Tool (BLAT) search using an online tool at https://genome.ucsc.edu/ (data not shown). Default settings were used.

Jeffers *et al*. (1990) initially cleaved murine and human genomic DNA with a restriction endonuclease and the digested samples were separated electrophoretically and transferred to a membrane. The Southern blot was then hybridised with the rat *Unr* probe (*v.s.*). The blot was then stripped and re-probed with a probe against the 5' region of murine *Nras*.

They found that there was one band for murine *Unr* that coincided with a band for *Nras* and inferred that the mouse haploid genome contains one *Unr* gene. For humans, they found two *UNR* bands, of which only one coincided with a band for *NRAS*. They inferred from this that the human genome has two *UNR* genes although they stated that they were unaware as to whether or not both genes could be translated.

Given that the method they used to probe for human *UNR* was indirect (*i.e.* they used part of the rat *Unr* gene), it was decided to investigate some seemingly unresolved issues surrounding their claims using public database searches.

Whilst there is little support for there being a second *UNR* gene in the human genome, the online BLAT tool at www.ensembl.org was used to carry out a search for sequences similar to the Jeffers probe using the GRCh38 version of the human genome. Default settings were used and results with E values above 10^{-10} were rejected. This provided support for there being a sequence

approximately as similar to that of Jeffers' probe on human chromosome 10 as the equivalent *UNR* sequence on chromosome 1 (Table 1.1).

Table 1.1: Regions of high similarity between regions of the human genome and Jeffers' probe containing part of the 3' untranslated region of the rat *Unr* transcript

Genomic Location	Orientation	Query Start	Query End	Length	Score	E-value	Identity (%)
1: 114716923-114717228	Reverse	1	309	310	485	2.60E-137	91.29
10: 58918950-58919064	Forward	1	115	115	168	6.40E-42	85.22
10: 58919073-58919117	Forward	127	171	45	77	1.90E-14	93.33
10: 58919130-58919193	Forward	187	250	64	104	1.80E-22	89.06
10: 58919201-58919252	Forward	257	309	53	84	1.50E-16	94.34

N.B. The sequence in yellow overlaps the *UNR* gene. The pink shading shows an almost continuous region of chromosome 10. It is not known why the method of calculating lengths in the first two rows appear to differ.

This observation suggests that Jeffers had mistaken the region on Chromosome 10 for a second *UNR* gene. It is worth noting that Jeffers was working before the age of the human genome project and readily available bioinformatics tools based upon it.

Database searches carried out within the Anderson lab drew attention to another interesting observation concerning *UNR* copy number (Ó Catnaigh, unpublished observations). The *UNR* transcript 'Homo sapiens cold shock domain containing E1 (CSDE1), transcript variant 4, mRNA' has 4330 nucleotides of which 18 were terminal adenines. The adenines were removed and the remaining nucleotides were BLATed against GRCh38 (Genomic sequence) using the tool at www.ensembl.org with default settings and an E-value cut off of 10^{-100}. This yielded 6 hits; four from the *UNR* genomic region and one region each from chromosome 7 and chromosome 5 (Table 1.2).

Table 1.2: Regions of high similarity between the mRNA sequence for human 'CSDE1, transcript variant 4' and human genomic DNA.

Genomic Location	Orientation	Query Start	Query End	Length	Score	E-value	Identity (%)
1: 114716914-114718217	Reverse	3009	4312	1304	2521	0.00E+00 (sic)	100.00
1: 114732603-114732819	Reverse	1495	1711	217	422	3.20E-118	100.00
1: 114739692-114739890	Reverse	661	859	199	386	1.50E-107	100.00
1: 114749821-114750208	Reverse	126	513	388	752	1.20E-217	100.00
5: 179082484-179083091	Reverse	2637	3249	614	1078	0.00E+00 (sic)	95.28
7: 88802635-88802955	Reverse	3272	3584	321	472	1.80E-133	88.16

N.B. The sequences in yellow overlap the *UNR* gene. The pink shading shows a 614 nucleotide sequence on chromosome 5 and the blue shading shows a 321 nucleotide stretch on chromosome 7.

The region on chromosome 7 overlapped the *ZNF804B* gene and the *UNR* transcript nucleotides that aligned to it lay within the same *UNR* exon highlighted in row 1 of Table 1.2 (i.e. Chromosome 1, 114716914-114718217). The region on chromosome 5 covered at least three *UNR* exons (Chromosome 1, 114718612-114718746, 114719579-114719746 and 114720539-114720718). This implied that it may have been a processed pseudogene of *UNR*. The BLAT tool stated that the region on chromosome 5 overlapped a gene, *ZNF354C*, and a known pseudogene, *RP11-281O15.7*. The pseudogene was found to be recorded as a:

"pseudogene similat (sic) to part of cold shock domain containing E1, RNA-binding CSDE1"

http://vega.sanger.ac.uk/Homo_sapiens/Gene/Summary?g=OTTHUMG00000163598;r=5:179082680-179083194;t=OTTHUMT00000374452

Therefore, as well as there being a genomic sequence similar to part of the *UNR* gene on chromosome 10, there also appears to be a processed pseudogene of *UNR* on chromosome 5.

1.1.6 Regulation of *UNR* transcription

The human *UNR* gene is transcriptionally repressed by the protooncoprotein MYC in conjunction with MAX (Mao et al., 2003). This is interesting when considered alongside the finding that UNR has a positive effect on the translation of the *MYC* transcript (Evans et al., 2003; *v.i.*).

1.2 The *UNR* transcript

1.2.1 The *Unr* transcript has a polarised distribution in rat muscle

The *Unr* transcript was shown to have a polarised distribution in rat muscle, where only around a fifth to a third as much *Unr* mRNA was found in the neuromuscular junction (NMJ) synaptic region in comparison to other regions of the cell. Interestingly, more than two times more was found in the NMJ synaptic region of denervated rat muscle cells when compared to innervated muscle cells. The underrepresentation of *Unr* transcript in the synaptic region was shown by both real time polymerase chain reaction (PCR) and microarray analysis (McGeachie, Koishi, Andrews, & McLennan, 2005).

1.2.2 *UNR* mRNA levels vary with cell type

The absolute amount of transcript found in mammalian cells varies greatly throughout the body. By means of example, it has been reported that there is approximately 7 times more *Unr* transcript present in the testes as compared to the liver (Boussadia et al., 1997). The transcript has been reported to be highly expressed in human MCF-7 cells (Sun et al., 2005).

1.2.3 The *UNR* transcript contains at least three different polyadenylation sites

It was claimed that cDNA formed from mouse, rat and human *Unr*/*UNR* transcripts each came in three different sizes – an observation that was explained as being a result of multiple 3' polyadenylation sites (Jeffers et al., 1990). Whilst this observation may have been valid, it is now known that different splice variants also exist (see below).

1.2.4 Major splice variants differ by presence or absence of exon 5

A 798-amino acid encoding open reading frame within *UNR* cDNA was suggested by Jeffers et al. (1990). The 798 amino acid protein has since been taken as the canonical UNR protein (www.uniprot.org).

It was later shown that UNR/Unr has a shorter splice variant that differs by the absence of exon 5 (Boussadia, Jacquemin-Sablon, & Dautry, 1993). Two serine residues within exon 5 were shown to be sites of phosphorylation within the UNR protein (Figure 1.2). There is also a lysine residue present within exon 5 that is predicted to have a 77% chance of being a site of ubiquitylation (Figure 1.3). This exon is responsible for encoding a series of 31 residues that lie between the first cold shock domain (see section 1.3.1) and the first pseudo cold shock domain (see section 1.3.1). It is reported that transcripts in which exon 5 is spliced out are generally more abundant than full-length transcripts except in brain tissue, where the two are equivalent in concentration (Boussadia et al., 1993).

The 798 and 767 amino acid-encoding transcripts are not the only transcript variants of *UNR*, however. A database search flagged up a variety of other transcripts, many of which being protein coding (Table 1.3). The exon structures, and protein lengths, of selected *UNR* transcripts are further presented in (Anderson & Ó Catnaigh, 2015).

Table 1.3: Reported/predicted *UNR* transcripts, with Ensembl IDs, and UNR protein lengths

Name	Transcript ID	bp	Protein	Biotype
CSDE1-201	ENST00000610726.4	4312	844aa	Protein coding
CSDE1-008	ENST00000438362.6	4167	844aa	Protein coding
CSDE1-002	ENST00000358528.8	4076	798aa	Protein coding
CSDE1-001	ENST00000339438.10	4006	767aa	Protein coding
CSDE1-004	ENST00000369530.5	3774	813aa	Protein coding
CSDE1-003	ENST00000261443.9	3228	767aa	Protein coding
CSDE1-009	ENST00000534699.5	2667	798aa	Protein coding
CSDE1-007	ENST00000530886.5	3459	668aa	Protein coding
CSDE1-010	ENST00000525878.5	626	46aa	Protein coding
CSDE1-016	ENST00000534389.2	601	94aa	Protein coding
CSDE1-014	ENST00000529046.5	571	136aa	Protein coding
CSDE1-015	ENST00000525132.1	521	36aa	Protein coding
CSDE1-013	ENST00000525970.5	503	7aa	Protein coding
CSDE1-005	ENST00000483407.1	741	No protein	Processed transcript
CSDE1-011	ENST00000524652.1	562	No protein	Processed transcript
CSDE1-012	ENST00000533818.5	552	No protein	Processed transcript
CSDE1-006	ENST00000483030.1	873	No protein	Retained intron
CSDE1-017	ENST00000530784.1	545	No protein	Retained intron

N.B. Downloaded on 24/08/16 from:
http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000009307;r=1:114716916-114750190;t=ENST00000438362

1.2.5 Regulation of the *UNR* transcript

One mechanism by which the *UNR* transcript level can be regulated has been mentioned above (see section 1.1.6). As MYC is a transcriptional repressor of *UNR* (Mao et al., 2003) and UNR is a translational enhancer of the *MYC* transcript (Evans et al., 2003), this allows UNR to limit its own transcription and, thereby, the number of *UNR* transcripts in the cell indirectly through modulation of the MYC protein level.

It has also been shown that Unr represses its own translation; the translational output of the *Unr* transcript increases around 3.6 fold when it has its UTR regions removed (Dormoy-Raclet, Markovits, Jacquemin-Sablon, & Jacquemin-Sablon, 2005). The authors demonstrated that the 3'-UTR destabilises the *Unr* transcript and also showed that the repressive effect of having its 5'-

UTR is reduced in *Unr* -/- cells. They provided evidence for their contention that Unr exerts an autoregulatory negative influence on its own translation through an interaction with an internal ribosome entry site (IRES) in its 5'-UTR. This shows that Unr can reduce its own translation by acting directly on its transcript as well as indirectly reducing its own transcription via increased translation of Myc. The polypyrimidine tract binding protein (PTB) also binds to and represses the capacity of the *UNR* IRES to undergo translational initiation events (Cornelis, Tinton, Schepens, Bruynooghe, & Beyaert, 2005).

1.2.6 Cell cycle dependent spike in cytoplasmic UNR

UNR and PTB stabilise an inactive conformation of the *UNR* IRES (Schepens et al., 2007). During early mitosis, the nuclear envelope breaks down and nuclear proteins become more free to interact with cytoplasmic proteins and mature mRNAs. The predominantly nuclear HNRNPC1/HNRNPC2 (hereafter termed hnRNP C1/C2), having been liberated from the nucleus, can bind to the *UNR* IRES. By doing so, they cause the dissociation of UNR and PTB and contort the transcript into a conformation conducive to translation initiation events occurring (Schepens et al., 2007). This does two things; it frees up UNR (and PTB) and it results in nascent UNR being translated. The increase in free UNR and PTB concentration then increases the probability of APAF1 being translated, potentially priming the cell for apoptosis during mitosis (Mitchell, Brown, Coldwell, Jackson, & Willis, 2001).

To add another level of complexity, both hnRNP C1/C2 and HDM2 also increase the translation rate of the X-linked inhibitor of apoptosis protein (XIAP) in a TP53-independent manner. This can then inhibit the activity of caspases such as caspase 3 (Figure 1.1; Holcik et al. 2003; Deveraux et al. 1997; Gu et al. 2009).

In relation to section 1.2.5, as well as UNR having a positive effect at the *MYC* IRES (Evans et al., 2003), hnRNP C1/C2 also have a positive influence on the *MYC* IRES, leading to increased translation at G2/M (Figure 1.1; Kim et al., 2003). PTB has also been shown to increase MYC translation through its IRES, at least in multiple myeloma (Figure 1.1; Cobbold et al. 2010).

	IRES trans-acting factor (ITAF)			
Transcript IRES	**UNR**	**PTB**	**hnRNP C1/C2**	**HDM2**
UNR				
APAF1				
XIAP				
MYC				

Key

positive effect

negative effect

Figure 1.1: Effect of selected IRES trans-acting factors (ITAFs) upon selected transcript IRES structures, see text for discussion.

1.3 The UNR protein

1.3.1 UNR contains 5 canonical cold shock domains

The UNR protein has five single-stranded nucleic acid-binding cold shock domains (Jacquemin-Sablon et al., 1994). The cold shock domain (CSD) is one of the most ancestral nucleic acid binding domains, being present in proteins across archaea (Giaquinto et al., 2007), eubacteria (Graumann & Marahiel, 1998) and eukaryotes (Mihailovich et al., 2010). The domain was noted to be present in bacterial proteins involved in the response to cold shock (reviewed in Schindelin, Marahiel, & Heinemann, 1993). There is no evidence that UNR is in any way related to responding to cold shock in mammalian cells, however. That said, whilst no functional explanation for the overrepresentation of *UNR* transcripts in testicular cells is currently available, it is interesting to consider the relative 'cold shock' experienced by all testicular cells in comparison to those within the internal organs.

The nucleic acid-binding domains of CSD-containing proteins allow them to function as RNA-chaperone proteins (Graumann & Marahiel, 1998). The five CSD architecture of UNR is unique (Brown & Jackson, 2004). UNR also possesses four regions that are closely related to cold shock domains but with mutations at critical residues that are predicted to render them non-functional. The location of these 'pseudo cold shock domains' (ψCSD), interspersed between adjacent CSDs (Figure 1.2), suggests that a series of duplication events led to the modern structure of the UNR protein.

1.3.2 UNR is well conserved among mammalian species and beyond

In general, the UNR protein is well conserved among metazoans. The protein alignment tool available at www.uniprot.org (accessed 28/07/14) was used with default settings to align the following 798 amino acid UNR homologs:

. H2N6A0 (Sumatran orang-utan),

. K7E2B4 (Gray short-tailed opossum),

. H9EMW2 (Rhesus macaque),

. P18395 (rat),

. Q91W50 (mouse),

. O75534 (human).

The alignment showed that all six proteins were identical at 756 of the 798 amino acids, equating to 95% identity (rounded to 2 significant figures). Reducing the number of species compared to human, macaque and orang-utan increased the identity to 797/798, rounding up to 100% (human and macaque UNR proteins were identical with Q713P in orang-utan). Even when a 799 residue homolog in the zebrafish was compared to the 798 residue human UNR, 622 residues were identical (78%).

1.3.3 UNR sequence is less well conserved in non-mammalian animal species

As stated previously, no homolog of UNR is believed to be present in *C. elegans*. Also, as of 24/08/2016, no *D. melanogaster* protein was termed CSDE1 in the Uniprot database (www.uniprot.org). There is, however, a wealth of literature on dUnr, published in large part by the Gebauer lab in Barcelona, Catalonia. Weighing against the general conservation apparent within the primary structure of UNR, *D. melanogaster* Unr, isoform A has some striking differences with respect to the human protein. The most salient difference is the presence of an N-terminal glutamine-rich region in dUnr across 63 residues, 49 are glutamine and 11 of the remaining 14 residues are histidine. The differences between documented UNR proteins in mammals and those in *D. melanogaster* imply a possible divergence in function. There is some evidence for this in the literature. For example, one function of dUnr that is not believed to be shared by human UNR concerns dosage compensation (reviewed briefly in Ray, Ó Catnaigh, &

Anderson, 2015). It is worth considering that the problem of gene dosage is different in different animal species and must, therefore, be addressed by different mechanisms (Payer & Lee, 2008).

One piece of earlier work that was potentially related to this work was an exploration that was carried out to identify dUnr-binding transcripts by both RIP-Seq and RIP-Chip (Mihailovich et al., 2012). That work identified transcripts encoding a wide variety of proteins including metabolic enzymes, transcription factors, signalling proteins and members of the ubiquitin-proteasome pathway. A number of housekeeping gene transcripts were also pulled down, including those encoding a variety of ribosomal proteins, and various isoforms of actin and tubulin subunits (Mihailovich et al., 2012).

Due to the known differences between mammalian and *D. melanogaster* UNR proteins, it was uncertain as to whether or not anything should be extrapolated from work on dUnr.

1.3.4 *D. melanogaster* Unr is an ortholog to human UNR

The database at www.uniprot.org contains information that shows human UNR and dUnr to be true orthologs.

To extract this information, the 'canonical' sequence for the human UNR protein (Uniprot identifier = O75534) was BLASTed against all arthropod proteins using the BLAST tool at www.uniprot.org with default settings on 24/08/2016. When only *D. melanogaster* proteins were considered, five sequences had E values less than 10^{-10}. Four were annotated as Unr and the other was not annotated to any gene. The E values for the proteins ranged from 20E-99 to 66E-129 (sic, values given as per www.uniprot.org). The best hit by E value was M9ND61 (Upstream of N-ras, isoform C). This was then BLASTed against all proteins in the human genome. The joint best hit by E value was O75534 (E value = 5.2E-129). The other joint top hit, 100% identical to O75534 at the protein level, was termed 'CSDE1, RNA binding, isoform CRA_a'. This means that human and *D. melanogaster* UNR proteins are reciprocal best hits and are strongly expected to have common ancestor. Interestingly, however, some other proteins also had highly significant alignments to M9ND61. These included the polyglutamine-containing ataxins, Ataxin 3 (best E value = 44E-24) and Ataxin 8 (best E value = 1.5E-24), both of which had over 70% identity over the aligned regions. The polyglutamine-rich N terminal region in most

versions of dUnr demonstrates a fundamental difference in primary sequence to that of human UNR.

1.3.5 The cold shock domains of *D. melanogaster* Unr retain stronger similarity to their human counterparts

Whereas dUnr has major differences in amino acid sequence from the protein found in humans, it retains stronger sequence identity within its cold shock domains. The first CSD (CSD1) is the most conserved; there is >70% identity between the human and *D. melanogaster* forms. It also exhibits the greatest similarity to bacterial cold shock domain containing proteins (Mihailovich et al., 2010). Different combinations of cold shock domains can be involved in mediating protein/transcript binding by any given UNR and any knock-on functions. Whilst different cold shock domains may function in unison to bind specific RNAs or proteins, a mere three residues within CSD1 were shown to mediate the binding of dUnr to either the dSxl protein or Dmel\msl-2 mRNA (Mihailovich et al., 2010). It should be noted, however, that the dUnr interaction with the Dmel\msl-2 transcript did not involve a phenylalanine residue and that deleterious effects on binding/function can be the result of other, unpredicted, effects of mutations.

Whilst strong conservation may be expected within protein domains that require a specific structure to retain their general function, it should nevertheless be noted that the reduced degree of conservation in sequences expected to bind single stranded nucleic acids implies that the *D. melanogaster* protein could have an entirely divergent set of mRNA binding partners. That implies that, whilst retaining the nucleic acid-binding function of a cold shock domain containing protein, dUnr may share little (if any) cellular functions with human UNR.

Due to the differences between dUnr and human UNR, both in sequence and function, it was decided not to consider dUnr any further in relation to this work.

1.3.6 RNA binding specificity of UNR

Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Tuerk & Gold, 1990) was used to predict consensus sequences of ribonucleotides for binding RNAs to human UNR (Triqueneaux, Velten, Franzon, Dautry, & Jacquemin-Sablon, 1999). This sequence was found to be AAGUA or AAC(G/A) located immediately downstream of 5-8 A/G nucleotides. UNR has been shown or suggested to have a variety of mRNA binding partners in the literature. Some of these originate from genes that exist throughout the metazoa. Others are more species-specific and show some diversity in UNR's mRNA binding partners has developed over evolutionary time.

At the point of writing (28th August 2016), the latest available update of the UTR database (http://utrdb.ba.itb.cnr.it/home/statistics) shows 656 potential UNR binding sites within 625 human 5'-UTRs and 3951 potential UNR binding sites within 3522 human 3'-UTRs. This uses the U0017 UNR binding site which is actually two sequences which closely resemble the reported SELEX sequences (Table 1.4). Given that this does not consider the potential for UNR binding within coding regions, it is clear that UNR could potentially directly interact with a substantial subset of human transcripts. Keeping this in mind, the identity of UNR-associated transcripts is an ongoing point of research for the Anderson lab.

Other reported Unr-binding RNA sequences include two closely apposed polypurine sequences within the major protein coding region determinant of instability (mCRD) from the *Fos* transcript (Chang et al., 2004; Table 1.4) and another purine rich sequence from the IRES within the coding region of the *CDK11* transcript that gives rise to the CDK11^{p58} form of the protein (Tinton, Schepens, Bruynooghe, Beyaert, & Cornelis, 2005; Table 1.4). Two observations should be noted concerning the *Fos* mCRD work. Firstly, whilst the reported regions believed to be involved in Unr binding were polypurine stretches, the actual sequences tested included some flanking pyrimidines. Secondly, the first sequence (termed 'PuSI') was tested directly for Unr binding whereas the support provided for the second sequence (PuSII) was indirect. Purine to pyrimidine mutations were placed into one of two polypurine sequences (PuSII and PuSIII) within a longer purine rich sequence and the resultant RNA molecule was tested for Unr binding. Mutations in PuSII caused the molecule to lose its affinity for Unr whereas mutations in PuSIII did not have this effect. It was from this analysis that the group inferred that PuSII is involved in binding Unr and that PuSIII is not (Chang et al., 2004).

Table 1.4: RNA sequences reported to bind UNR/Unr

Nucleotide Sequence	Source	Reference
*AAGUA	SELEX	**Triqueneaux *et al*., 1999**
*AAGUG	SELEX	**Triqueneaux *et al*., 1999**
*AACG	SELEX	**Triqueneaux et al., 1999**
**RRRRRAAGUAR	UTRsite	http://utrsite.ba.itb.cnr.it
**RRRRRRRRAACRRR	UTRsite	http://utrsite.ba.itb.cnr.it
AGAAGAAGAAGAGAAAAGGAGAA	PuS I from *Fos* mCRD region	**Chang *et al*., 2004**
GAAGGGAAAGGAA	PuS II from *Fos* mCRD region	**Chang *et al*., 2004**
GAAGAAGUAAA	IRES from *CDK11* transcript	**Tinton *et al*., 2005**

* This is found downstream of a purine stretch. ** R = A or G (i.e. a purine).

1.3.7 Post-translational modifications 1 - phosphorylation

Three serine residues have been shown to be phosphorylated in UNR (S116, S123 and S514) (Figure 1.2; Dephoure et al., 2008; Olsen et al., 2010). Interestingly, S116 and S123 were found to be phosphorylated during mitosis but not during G1 (Dephoure et al., 2008). This raises an interesting question about a potential isoform-specific role for UNR in mitosis, given that most UNR transcripts were shown to lack exon 5 (Boussadia et al., 1993).

1.3.8 Post-translational modifications 2 – acetylation

K81 is reported to be susceptible to acetylation (Figure 1.2; Choudhary et al., 2009).

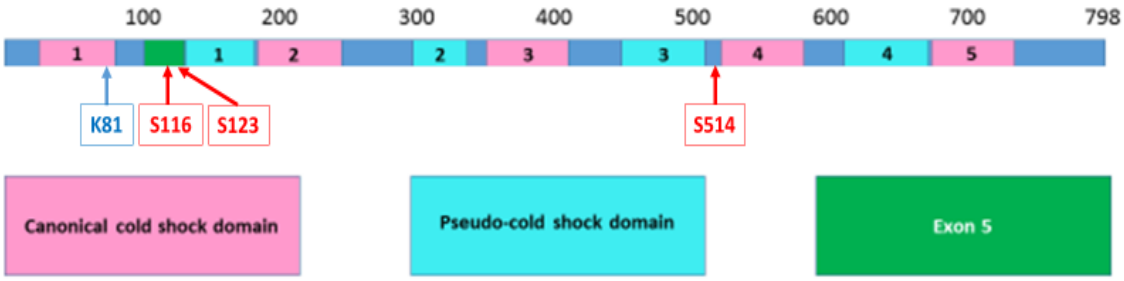


Figure 1.2: Schematic diagram of full length UNR protein showing the approximate location of its five canonical cold shock domains (pink), its four pseudo-cold shock domains (turquoise), exon 5 that is missing in the most common forms of the protein (green; see text for details). The blue arrow points to lysine 81, reported to be a site of acetylation. The red arrows point to three serines (serine 116, serine 123 and serine 514) noted to be phosphorylated under specific conditions (see text for details). N.B. K81 lies within CSD1 and both S116 and S123 lie within exon 5.

1.3.9 Post-translational modifications 3 – other potential modifications

There are multiple other potential sites of phosphorylation, ubiquitylation, SUMOylation, etc. that are easily obtained from searching online databases (Figure 1.3; condensed and reproduced in Anderson & Ó Catnaigh (2015).

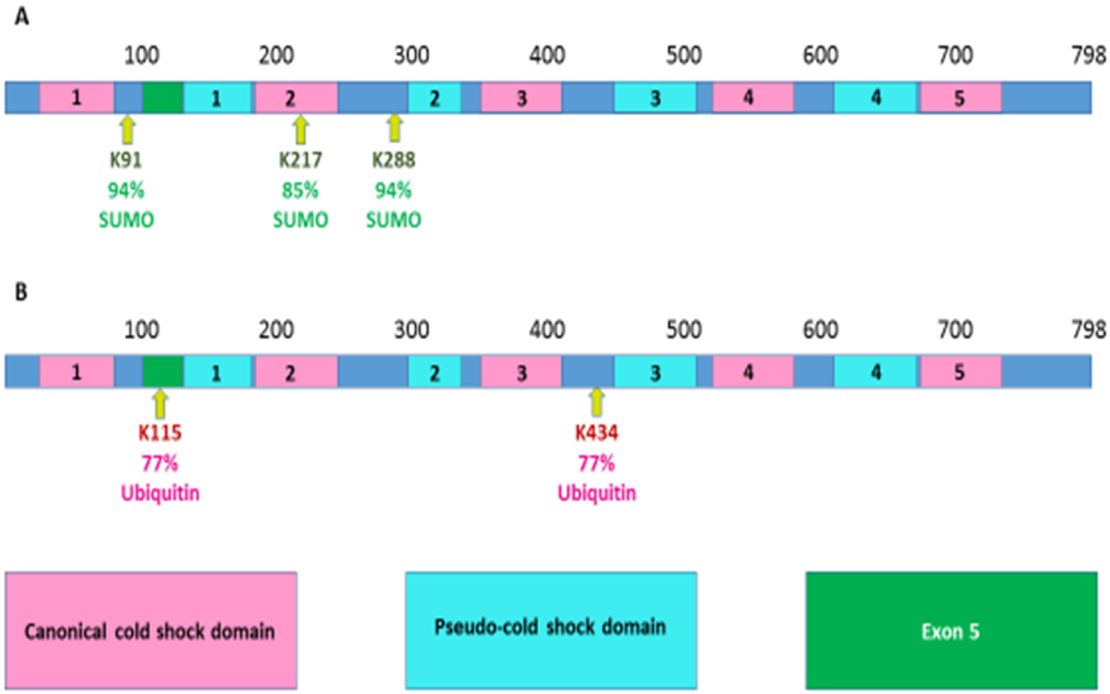


Figure 1.3: Schematic diagram of full length UNR protein showing the approximate location of its five canonical cold shock domains (pink), its four pseudo-cold shock domains (turquoise), exon 5 that is missing in the most common forms of the protein (green; see text for details). The figure shows: A) the top three most probable sites of potential SUMOylation within full length UNR with associated probabilities and, B) the top two most probable sites of potential Ubiquitylation within full length UNR with associated probabilities. N.B. the K115 site of potential ubiquitylation is within Exon 5.

1.4 UNR as a modulator of translational initiation

1.4.1 Cap-dependent initiation

The processes of transcription and post-transcriptional modification generate mRNA species that are exported from the nucleus. Typical eukaryotic transcripts have a 5' cap and a 3' poly(A) tail. A variety of canonical initiation factors contort mRNA into a circularised structure that is conducive for efficient translation. Briefly, EIF3 binds to the small ribosomal subunit and then to EIF2 that has bound an initiation methionyl-tRNA. Then EIF3 binds to EIF4G which is part of the EIF4F complex, together with EIF4A (4A) and the cap-binding EIF4E, to which a capped transcript is bound. The helicase 4A helps to smoothen out any complex RNA structures that could otherwise prevent or impede translation. Multiple copies of poly(A) binding protein (PABP) are

bound to the 3' poly(A) tail of the transcript and PABP also binds 4G, thereby generating a physical link between the 5' and 3' ends of the transcript. The 5' end of the transcript is then scanned for an initiation codon (usually AUG). Once this is found, inhibitors of large and small ribosomal subunit binding are removed and a functional ribosome is formed on the transcript. The next stage in translation, elongation, can then proceed. These steps are reviewed in more detail in (Jackson, Hellen, & Pestova, 2010).

1.4.2 UNR as a global regulator of translation

Work within the lab has led to the suggestion that UNR may be a global regulator of translation (Ray & Anderson, 2016). This work showed that recombinant wild type UNR stimulates the *in vitro* translation of *Renilla* luciferase mRNA around 5 fold when the transcript is capped and polyadenylated. The increase was about 4 fold when polyadenylation was lost, almost 40 fold when the cap was lost and approaching 200 fold when both the cap and polyadenylation were lost. The experiment also involved the use of recombinant mutant UNR proteins with an essential phenylalanine from each cold shock domain being mutated. UNR with a mutation in cold shock domains 1, 3 or 5 generally provided a similar or better increase in translation to wild type UNR across all conditions. UNR with a mutation in cold shock domain 4 and, more so, cold shock domain 2 showed a reduced increase in translation relative to wild type UNR. The paper went on to show that UNR interacts with PABP *in vitro* but that the interaction is reduced by 71% with the CSD2 mutant UNR, by 76% with the CSD4 mutant UNR and by 21% when RNase A is added to the rabbit reticulocyte lysate. This implied that UNR interacts with PABP, which was already known (Chang et al., 2004), in large part via cold shock domains 2 and 4. It also implied that the interaction was stabilised by the presence of RNA. Data from one of the cell types used during the course of the present study (HeLa) then showed that knocking down UNR reduced the amount of 4G pulled down by PABP by 80%. Interestingly, the amount of UNR and PABP pulled down with 4G in the presence of RNase ONE both dropped by 40% following UNR knockdown where the knockdown was much greater than 40%. This led to the suggestion that UNR binds Pabp directly and EIF4G via PABP and that UNR stabilises the interaction between PABP and EIF4G. Finally, it was shown in two cell lines used in the current study (HeLa and U2OS) that knocking down UNR resulted in a reduction of ^{35}S-methionine incorporation of around a third (U2OS) or around a half (HeLa). This implied that the addition of UNR (i.e. not knocking it down) increases global translation in human cell lines as well as *in vitro* (Ray & Anderson, 2016).

1.4.3 Cap-independent initiation via IRES structures

Under certain pathological and non-pathological conditions, such as infection with certain viruses or during mitosis, cap-dependent initiation events are prevented or reduced in frequency. A widely held view states that a proportion of transcripts contain complex folded structures termed internal ribosome entry sites (IRESes) within their 5' untranslated region (5'UTR) or within their coding sequences (reviewed in Jackson, 2013). These structures can bind to IRES trans-acting factors (ITAFs) that can modulate their activity. These ITAFs, in conjunction with other canonical and non-canonical initiation factors, can contort transcripts into conformations that are conducive for translational initiation to occur, effectively side-stepping the need for the cap and cap-binding proteins. The process is less rigidly choreographed in comparison with cap-dependent initiation events as the identity of the ITAFs shows some transcript selectivity. This is to be expected as, unlike 5' caps and PABP proteins, IRES structures are variable in topological structure (Jackson, 2013).

IRES structures are common in RNA viral genomes as they offer a mechanism by which the virus can dispense with the need to obtain a cap to be able to undertake efficient translation initiation events (Jackson, 2013). It should be noted that the concept of cellular IRESes, whilst currently the prevailing consensus view within the field, is not universally accepted. One alternative, though not necessarily contradictory, hypothesis states that cap independent scanning of transcripts occurs with the aid of cap-independent translational enhancers (CITEs) that can bind to either the 3' or the 5' regions of a transcript (for a review of this, see Shatsky, Dmitriev, Terenin, & Andreev, 2010). The remainder of this thesis will make the assumption that cellular IRESes do exist and that UNR is an ITAF for a selection of both cellular and viral IRES structures.

1.4.4 UNR as an ITAF

UNR has been reported to modulate initiation events through a number of IRES structures, as described below. As an example, UNR was identified as being important for the translation of the rhinovirus RNA through its IRES and was later shown to be used by poliovirus for the same role (Boussadia et al., 2003; Hunt, Hsuan, Totty, & Jackson, 1999). All five CSDs are required for the efficient translation of the human rhinovirus RNA, however. CSD1 and CSD2 were suggested to be most important in that role, as UNR proteins in which conserved phenylalanine residues,

thought to be important in RNA binding, were mutated to alanine in either of those CSDs totally abrogated the stimulatory effect of UNR. Similar mutations in the other CSDs merely reduced the effect (Brown & Jackson, 2004). It is worth considering this finding in conjunction with the finding that CSD2 was important in the interaction between UNR and PABP (Ray & Anderson, 2016). The same authors later showed that UNR has two independent binding sites on the rhinovirus IRES (Anderson, Hunt, & Jackson, 2007).

Like UNR, PTB can also be usurped by viruses for the purposes of enhancing their capacity for translation initiation. For example, PTB was shown to enhance translation through the foot and mouth virus IRES and multiple copies of PTB were shown to bind to the IRES of encephalomyocarditis virus (Kafasla et al., 2009; Niepmann, 1996). It is also involved, with UNR, in translation initiation at the poliovirus and human rhinovirus IRES (Boussadia et al., 2003). Indeed, UNR was only discovered to be necessary for activation of the rhinovirus IRES after PTB had previously been shown to be necessary but not sufficient (Hunt and Jackson, 1999).

There are a number of reported cases of UNR functioning as an enhancer ITAF at cellular IRESes. Several of these, such as the transcripts encoding *APAF1* and *CDK11* will be discussed below (v.i.). UNR has also been reported to bind to an IRES region within its own transcript, where it acts to stabilise an inactive conformation and, thereby, reduce the probability of IRES-mediated initiation events occurring (v.i.).

The main point here is that UNR binds to different transcripts in conjunction with different protein partners (reviewed in Ray et al. 2015). This is important because it provides the cell with a mechanism to carefully control UNR function by controlling the levels and localisation of a variety of other proteins aside from UNR itself.

1.4.5 Non-ITAF molecular functions of UNR in translational regulation

As well as stimulating or repressing translation through IRES structures, Unr has also been shown to function in the co-translational regulation of *the Fos* transcript. As discussed in section 1.5.3, Unr can stabilise the *Fos* transcript until it is translated, at which time Unr dissociates from the transcript, leading to Ccr4-dependent deadenylation and transcript degradation (*v.i.*). This shows Unr acting as both a transcript stabilising factor and a modulator of transcript degradation (Chang et al., 2004; Grosset et al., 2000).

UNR has been shown to interact with an A-rich autoregulatory region within the 5'-UTR of the mRNA for poly-A binding protein (*PABP*) and inhibit its translation (Patel, Ma, & Bag, 2005). Drawing on an earlier finding that Unr is unable to bind poly(A)-RNA (Chang et al., 2004), the assumption was made that UNR required PABP to associate with the region.

The negative feedback phenomenon by which parathyroid hormone (PTH) levels are controlled by, and, in turn, control the systemic calcium level was observed several decades ago (Walton, 1977). PTH/Pth levels change in response to changes in the concentration of such factors as vitamin D, calcium and phosphate in PTH/Pth-producing cells, the latter two exerting their effect post-transcriptionally (Kumar & Thompson, 2011; Silver, Yalcindag, Sela-Brown, Kilav, & Naveh-Many, 1999). It was reported that hypocalcaemia stabilised the *Pth* transcript whereas hypophosphataemia was associated with its rapid degradation (Moallem, Kilav, Silver, & Naveh-Many, 1998). PKA was later suggested to play a role in the control of PTH levels in relation to low calcium and PKC was postulated to be involved in basal expression levels (Moallem, Silver, & Naveh-Many, 1995).

UV cross-linking of the 3'UTR of *Pth* mRNA showed that three proteins interacted with its AU-rich element in rat parathyroid tissue (Moallem et al., 1998). When the extracts came from animals subjected to a low calcium diet immediately prior to being sacrificed, there was more overall binding and the converse was true with low phosphate-fed animals (Moallem et al., 1998). The same three proteins were observed when extracts from certain other tissues were used but an increase in binding with low calcium and decrease with low phosphate was not

apparent. More overall binding was noticed with brain tissue extract in comparison to other tissue extracts examined (Moallem et al., 1998).

Further research by the same group showed that one of the three proteins that bound to the extreme 3'UTR of the *Pth* mRNA was the AU-rich binding factor 1 (AUF1) (Sela-Brown, Silver, Brewer, & Naveh-Many, 2000). Interestingly, the group went on to show that UNR was another of the three proteins and that its cellular concentration was positively correlated with Pth expression (Dinur, Kilav, Sela-Brown, Jacquemin-Sablon, & Naveh-Many, 2006). Possibly coincidentally, it was noted elsewhere that the predominant *UNR* transcript in most adult rat tissues lacked exon 5 except brain tissue where transcripts encoding full-length UNR were almost as abundant (Boussadia et al., 1993). Bringing the developmental role of UNR into consideration here, it is worth noting that the rate at which exon skipping occurs increases through development and into adulthood in a range of murine tissues (López-Fernández, López-Alañón, & del Mazo, 1995).

As research developed over time, the control of *PTH*/*Pth* mRNA stability and output came to be thought to obey the following model (Nechama, Peng, et al., 2009; Nechama, Uchida, Mor Yosef-Levi, Silver, & Naveh-Many, 2009):

UNR associates with an AU rich element (ARE) in the 3'-UTR of the *PTH* transcript under low calcium levels. This stabilises the transcript and increases PTH expression. Under low phosphate levels, UNR and AUF1 dissociate from the transcript and KH-type splicing regulatory protein (KSRP, also known as FBP2) binds. This, in turn, recruits the exosome and the endonuclease, polysome ribonuclease 1 (PMR1) and leads to the cleavage of the transcript and a subsequent reduction in the PTH concentration. The stability is further mediated by the peptidyl prolyl isomerase PIN1. Under basal conditions of calcium and phosphate, both KSRP and the UNR/AUF1 complex can bind to the transcript. Under these conditions, active PIN1 causes KSRP1 to be active (dephosphorylated) and recruit the exosome. Inactive PIN1 or the absence of PIN1 results in KSRP1 also being inactive (phosphorylated), thereby resulting in it being unable to bind to the transcript and this, in turn, results in transcript stability (Figure 1.4).

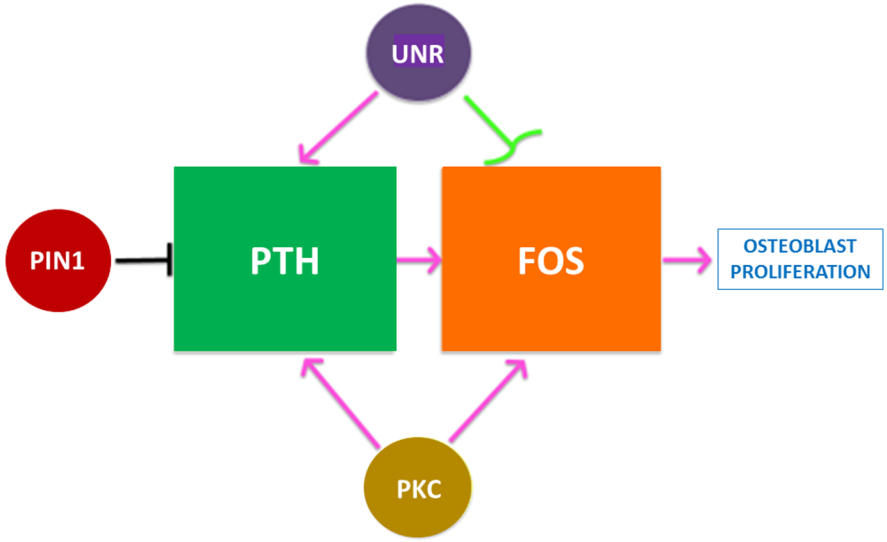


Figure 1.4: Model to show effects of PIN1, UNR and PKC on osteoblast proliferation. Data is, of necessity, from different cell types. Pink arrows show a positive influence on expression. Black blunt arrow shows strong repression on expression. The green curly arrow shows that UNR promotes both transcript stability and transcript degradation.

As an interesting aside, in its guise as FBP2, KSRP1 has a negative effect on the translation of the Enterovirus 71 picornavirus RNA (Lin, Li, & Shih, 2009). As stated above, it is known that UNR promotes the translation of the RNA of other picornaviruses. Indeed, UNR was shown to interact with the 5'UTR of the Enterovirus 71 RNA (Lin et al., 2008). It was also recently demonstrated that AUF1 associates with Enterovirus 71 IRES. Unlike with the PTH transcript, however, AUF1 has a negative effect on its translation (Lin, Li, & Brewer, 2014).

1.5 Suggested functional roles for UNR

1.5.1 Possible role in apoptosis

UNR has been implicated in modulating the translation of the apoptosome scaffold protein Apoptotic protease activating factor 1 (APAF1) through an internal ribosome entry site (IRES) in the 5' untranslated region of its transcript (Mitchell et al., 2001). Translational activity is stimulated in the presence of both UNR and the polypyrimidine tract binding protein (PTB). Later work by the same group showed that the *APAF1* IRES is contorted into a conformation conducive

to translational initiation upon binding to the both UNR and PTB (Mitchell, Spriggs, Coldwell, Jackson, & Willis, 2003).

APAF1 is an essential component of the intrinsic pathway of apoptosis. Following upstream signalling leading to mitochondrial outer membrane permeablisation (MOMP), cytochrome c is free to enter the cytoplasm. The binding of cytochrome c causes a conformational change in APAF1 that promotes its oligomerisation into the apoptosome backbone. APAF1 is not an inert scaffold; the oligomer possesses an active proteolytic domain that can cleave procaspase 9 to form active caspase 9. Caspase 9 is then free to activate the effector caspase, Caspase 3 (reviewed in Bratton & Salvesen, 2010). A reduction in the *APAF1* transcript level and the APAF1 protein levels in melanoma is strongly correlated with metastasis, possibly due to promoter methylation and/or excessive histone deacetylase activity (Soengas et al., 2001). In keeping with the idea of histone acetylation, trichostatin A was shown to enhance transcription of the *APAF1* gene. This also increased the transcription of E2f1 and Tp53, two transcription factors involved in *Apaf1* transcription (Wallace & Cotter, 2009). Interestingly, *APAF1* translation is believed to be exclusively initiated through the IRES ultimately discovered to be controlled by UNR (Coldwell, Mitchell, Stoneley, MacFarlane, & Willis, 2000).

It is important here to consider that APAF1 is not exclusively involved in the control of apoptosis. For example, the *APAF1* promoter was found to be methylated in various testicular germ cell cancers (TGCCs) but the protein was not linked to apoptosis in these cells (Behjati et al., 2011; Christoph et al., 2006). The amount of APAF1 protein expressed in different TGCCs was, however, found to correlate inversely with the degree of differentiation of the cells within the cancer. (Behjati et al., 2011). It was further shown that APAF1 translocates to the nucleus following treatment with cisplatin in A549 cells and the knocking down of APAF1 abrogates the capacity of cisplatin to cause cell cycle arrest in the same cell line (Zermati et al., 2007). The same study showed that APAF1 knock-down can also reduce the amount of CHK1 phosphorylation in response to γ-irradiation in the same TP53-deficient cells. To summarise, APAF1 is both up- and down-stream of TP53 and possibly plays a role in the DNA damage response as well as apoptosis. Of interest to the work at hand, whatever multiple functions APAF1 has in the cell, UNR is required for the efficient translation of its transcript.

As well as regulating the translation of the *APAF1* transcript, further work implicates UNR as a potential master regulator of apoptosis. For example, whilst *Unr* null murine embryonic stem

(ES) cells were shown to exhibit no deleterious effects in terms of cellular proliferation, the ability of such cells to undergo apoptosis in response to ionising radiation (IR) was reduced. Furthermore, expression of transcripts for the apoptosis-mediating proteins Tp53, caspase-3 and Gadd45γ were all reduced (Dormoy-Raclet et al., 2007). The same work also found that siRNA-knockdown of UNR in HuH7 cells doubled the proportion of apoptotic cells in response to IR and slightly increased the proportion in untreated cells. Barring the possibility that the increase in apoptosis was caused by off-target effects of the siRNA or some other mutation within the HuH7 cancer cell line, the implication is that UNR may have tissue-specific effects. A larger list of differentially expressed transcripts between *Unr*-wild type and *Unr*-null ES cells, with or without IR stress, had been reported previously by the same author (Dormoy-Raclet, 2005).

UNR is also known to modulate the translation of proteins involved in functions not directly related to apoptosis, such as cytokinesis (see section 1.5.2) and cellular proliferation (see section 1.5.3). This raises the possibility that UNR may act as a more general master regulator of translation. It should also be noted, however, that failure to undergo cytokinesis and/or uncontrolled proliferation are likely to cause apoptosis and UNR may be acting as an apoptotic switch that integrates and assesses information on cellular viability and becomes activated when cellular homeostasis is compromised in some way.

1.5.2 Suggested role for UNR in mitosis

It has been suggested that UNR plays an important role in regulating the translation of proteins during mitosis. In this model, UNR binds to, and represses translation of, its own transcript in conjunction with polypyrimidine tract-binding protein (PTB). Following breakdown of the nuclear envelope in early mitosis, hnRNP C1/C2 can come into contact with the cytoplasmic translation machinery. These outcompete UNR for binding on the UNR transcript, leading to the detachment of UNR and PTB. This drives the translation of UNR from its IRES (Schepens et al., 2007). Logically, having a cell cycle dependent expression pattern greatly increases the probability of a given protein being involved in some form of cell cycle dependent processes itself. The liberated and nascent UNR proteins can then promote the translation of cell cycle regulator CDK11^{p58} from an IRES in the coding region of its transcript during G2-M (Schepens et al., 2007; Tinton et al., 2005).

Transcription of *Cdk11* results in a transcript that can be translated to form a constitutive 110kDa kinase that functions within a multi-protein complex involved in transcriptional regulation and the splicing of pre-mRNA species (Li, Inoue, Lahti, & Kidd, 2004). IRES-mediated translation results in the generation of a smaller 58kDa product that functions in the regulation of cytokinesis. Unr has been shown to stimulate IRES-mediated translation of Cdk11 between G2 and M phase. Whereas humans have two *CDK11* genes, mice have a single copy; the loss of which results in death of the blastocyst prior to E4. It was suggested that lethality is caused by excessive apoptosis in response to mitotic arrest (Li et al., 2004). Interestingly, during the progression of extrinsic pathway-instigated apoptosis, $CDK11^{p110}$ is cleaved by caspase 3 and the N-terminal non-kinase domain-containing fragment translocates from the nucleus to the mitochondrion where it dissipates the potential across mitochondrial membrane and induces the release of cytochrome c into the cytoplasm (Feng, Ariza, Goulet, Shi, & Nelson, 2005).

1.5.3 UNR and the proto-oncogenes *MYC* and *Fos*

UNR regulation of the translation of some common proto-oncogenes has been documented in the literature. For example, it has been suggested that UNR can increase the translation of *MYC* mRNA through an IRES in conjunction with UNR-interacting protein (UNRIP) and a variety of other non-canonical initiation factors (Evans et al., 2003).

Similar proteins to those noted to enhance the translation of *MYC* through its IRES, including UNR, were also noted to form a complex on the Major Protein Coding Region Determinant of Instability (mCRD) of *Fos* mRNA (Grosset et al., 2000). As the complex also included Pabp, it was able to link the mCRD to the poly(A)-tail. Interestingly, the regulation of the transcript by Ccr4-mediated deadenylation was linked to the rate of translation; non-translating transcripts were not degraded. Reducing cellular Unr with siRNA reduced the rate of degradation (Chang et al., 2004).

1.5.4 Suggested essential role in embryonic development

From coitus and subsequent conception, mouse concepti start to undergo a series of well-choreographed cell divisions (reviewed in Kojima, Tam, & Tam, 2014). Briefly, the earliest cell divisions give rise to a solid ball of around 20 cells, termed the morula. Within this ball of cells, the first major cell fate decision is taken such that the outer cells differentiate into trophectoderm and the inner cells develop into inner cell mass cell. From this point, the cells of conceptus develop a clear morphological polarity as the blastocyst cavity starts to appear within the ball. This pushes the inner cell mass to one end and the conceptus, still surrounded by the *zona pellucida*, is termed the blastocyst and its cells become termed blastomeres. The blastomeres of the inner cell mass then make a second major cell fate decision, with some becoming primitive endoderm (predominantly extraembryonic) and the remainder becoming pluripotent epiblast. As it is relevant to later discussion, a simplified model of this process shall be discussed briefly at the molecular level.

Undifferentiated inner mass cells express both Gata6, a marker for primitive endoderm, and Nanog, a pluripotency marker. These transcription factors exert a mutually repressive downstream effect on each other (Bessonnard et al., 2014). Nanog is involved in promoting the expression of the paracrine fibroblast growth factor receptor ligand, Fgf4 (Frankenberg et al., 2011). This can diffuse to Fgfr2-expressing cells and instigate a MAPK cascade in cells destined to differentiate into primitive endoderm, resulting in the activation of Erk by phosphorylation. Phosphorylated Erk then reinforces the asymmetry in Gata6 level by promoting its further expression (Frankenberg et al., 2011). Reinforcing this point, pharmacological inhibition of Erk resulted in a reduction in Gata6 expression (Hamilton et al., 2014). Activation of the Fgfr2 was further shown to downregulate *Nanog* at the transcriptional level (Santostefano, Hamazaki, Pardo, Kladde, & Terada, 2012). Mouse embryonic stem cells lacking *Nanog* were shown to express a high level of *Gata6* and to differentiate into a primitive endoderm-like state that was claimed to lack Oct4 and other markers of mature primitive endoderm (Mitsui et al., 2003). Overexpression of *Gata6*/*GATA6* was also shown to cause differentiation into primitive endoderm independent of Fgf signalling (Wamaitha et al., 2015). Conversely, Gata6-null embryos completely lack a primitive endoderm (Schrode, Saiz, Di Talia, & Hadjantonakis, 2014). A simplified model of the cell fate decision of inner cell mass cells to become epiblast or primitive endoderm is provided in Figure 1.5 (*v.i.*).

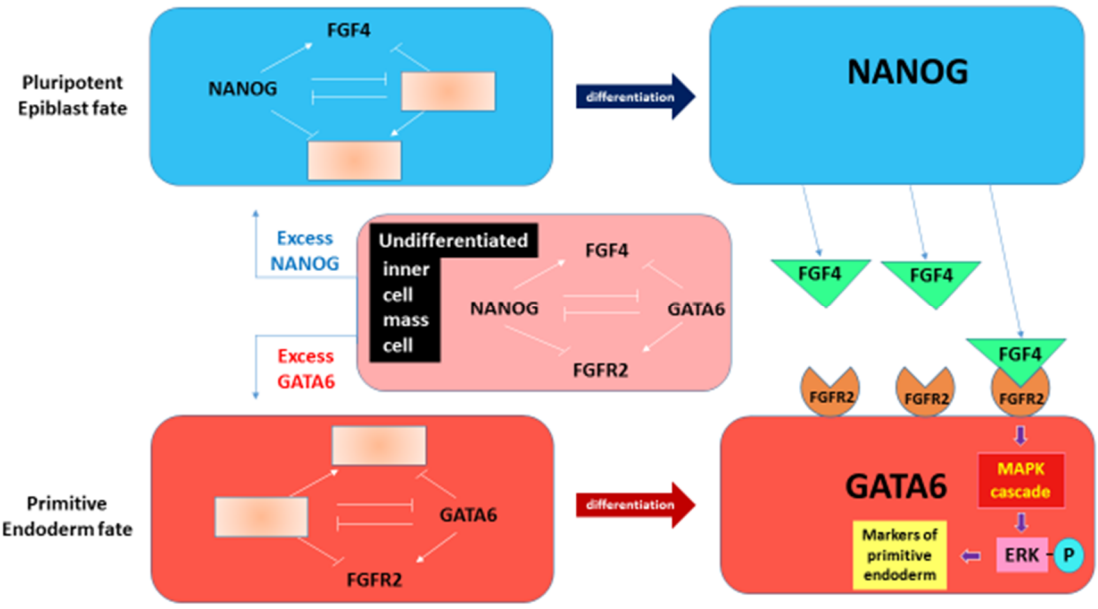


Figure 1.5: Variations in the amount of NANOG and GATA6 in undifferentiated inner cell mass cells (pink) are responsible for a cell fate decision. Excess NANOG represses GATA6 expression and leads to the retention of stemness and differentiation into pluripotent epiblast (blue). Conversely, excess GATA6 represses NANOG expression and results in cells differentiating into primitive endoderm (red). The latter is accompanied by the activation of a mitogen activated protein kinase (MAPK) cascade by the fibroblast growth factor receptor 2 (FGFR2), leading to the phosphorylation and activation of an extracellular signal regulated kinase (ERK). Phosphorylated ERK potentiates the expression of GATA6.

It was noted that Unr is associated with preventing epiblast cells from differentiating inappropriately into a primitive endoderm-like lineage that expressed pluripotency markers such as Nanog and Oct4 (Elatmani et al., 2011). Here, the authors used two techniques to obtain murine embryonic stem cells (ESCs) that were deficient in *Unr*. One set came from selecting *Unr* null early embryos formed by the crossing of two mice heterozygous for a deletion in the *Unr* promoter, as discussed previously. A second method involved stably knocking down *Unr* by means of shRNA. They found that cells deficient in *Unr* adopted a primitive endoderm-like phenotype in which Gata6 was heavily expressed. When they reintroduced a tagged Unr via a retroviral construct, they found that Unr was expressed although at a much lower level than in wild type cells. A tagged retroviral insert alone had no effect on Unr levels. They noted that even a very low level of Unr expression was sufficient to reduce primitive endoderm markers such as Gata6 and Gata4 by more than half. Whilst showing that shRNA-mediated ablation of *Unr* could cause differentiation into a primitive endoderm-like state, the authors also showed that both wild type and *Unr*-null ESC retained the capacity to generate teratomas in immunodeficient

mice. The removal of Unr, therefore, does not affect the differentiation potential of ESC *per se*. Earlier work in the same lab had shown that the amount of *Unr* transcription in a given cell type can affect the transcriptional output of the *Nras* gene (Boussadia et al., 1997). They therefore explored the possibility of Nras having affected the differentiation pathway. This was of interest as a signalling pathway had previously been elucidated between the activation of fibroblast growth factor receptors and the activation of Erk via Ras (Kouhara et al., 1997). They did this by looking at the absolute amounts of *Nras* transcripts in wild type cell, those with *Unr* knocked out and those with *Unr* stably knocked down with or without the retroviral *Unr* insertion. They found that the amount of *Nras* transcripts was 2-3 times greater in *Unr* knockouts compared to the other three conditions (in which *Nras* transcripts levels were similar). This led them to conclude that the differentiation or otherwise of Unr depleted cells into primitive endoderm was independent of the *Nras* mRNA concentration. They then examined the protein and transcript concentrations of Nanog/*Nanog* between Unr-wild type and Unr-null ESCs. They noted that there was no significant difference and inferred that Unr must act downstream of Nanog. The presence of Nanog highlighted two things; firstly, Nanog was partially located in the cytoplasm contrary to the received wisdom as to its localisation in cells. Secondly, the joint expression of Nanog and Gata6 implied that the cells were not true primitive endoderm. The authors claim that the state is an early stage in the differentiation process. They show that the removal of Unr expression results in the stabilisation of *Gata6* mRNA as measured by transcript half-life. They further provide evidence that suggests that Unr may act to destabilise the *Gata6* transcript through an interaction with its 3'-UTR.

Given the link between UNR, GATA6 and cell fate decisions, it is interesting to consider that GATA6-FGFR2 over-expression is associated with the cancer phenotype, e.g. in oesophageal adenocarcinoma (Lin et al., 2012). These authors showed a dependence of cancer cells on GATA6 expression such that, when GATA6 was depleted, the cells underwent apoptosis. Conversely, it has been shown that Gata6 overexpression is associated with cellular quiescence (Nagata et al., 2000). Therefore, whilst UNR may modulate the expression levels of GATA6, the functional response could be context dependent.

1.5.5 UNR and Diamond Blackfan Anaemia

A new line of investigation into the function of UNR involves a link to the ribosomopathy-related blood disorder, Diamond Blackfan Anaemia (DBA) (Da Costa et al., 2010). The UNR protein level is reduced in erythroblasts from patients with DBA and it was shown that knocking down *Unr* in *Tp53*-null mouse erythroblasts reduced both proliferation and differentiation (Horos et al., 2012).

1.5.6 UNR and the migration of precerebellar neurons

It has been suggested that Unr has a developmental role in the correct migration of precerebellar neurons in the developing hindbrain (Kobayashi, Kawauchi, Hashimoto, Ogata, & Murakami, 2013). They showed that the *in utero* knocking down of *Unr* in these cells resulted in significant numbers failing to reach their target region in 19/19 brains tested, whereas this was the case in none of eight brains containing control-transfected cells (Kobayashi et al., 2013). They suggest that Unr may be involved in the post-transcriptional regulation of a group of mRNAs involved in the migration process.

1.5.7 UNR and autism

It has been suggested that UNR may play a role in autism with a gene that lies in close proximity to it in the human genome, TRIM33 (Xia et al., 2013). Links between UNR and TRIM33 are discussed in more depth later (see section 4.3.6).

1.5.8 UNR and Alzheimer's disease

Amyloid precursor protein (APP) is a transmembrane protein that has been implicated in the pathogenesis of Alzheimer's disease. Briefly, the protein is processed by one of two proteases (α- or β-secretase) that differ in their site of cleavage. The processed protein is then acted on by another protease, γ-secretase. This results in a variety of cleavage products; one of those formed

by the action of β- and γ-secretase is the precursor of the amyloid plaques that are believed to be responsible for the development of Alzheimer's disease (reviewed in Renner, 2014). It is believed that the protein ADAM10 (a disintegrin and metalloproteinase 10) acts as an α-secretase. UNR was shown to interact with the 5'UTR of the *ADAM10* transcript. It was shown to stabilise the transcript without repressing translation (Renner, 2014). Linked to this, *UNR* knockdown resulted in reduced levels of ADAM10, thereby potentially increasing the risk of more plaque precursor proteins being formed. Interestingly, that outcome was not observed although it was suggested that this may have been due to the cell line being used (Renner, 2014). Paradoxically, although *UNR* knockdown did not increase the amount of plaque precursor, overexpression of UNR did increase it. It was postulated that UNR may increase the translation of the BACE1 β-secretase through a hypothetical interaction with the 3'UTR of its transcript (Renner, 2014).

In other work, *UNR* expression was shown to be 50% lower in Alzheimer's Disease-affected hippocampi relative to control samples (Acquaah-Mensah, Agu, Khan, & Gardner, 2015). It is interesting to consider this finding in light of the ADAM10 work. A reduction in UNR levels could be expected to destabilise the ADAM10 transcript but, if Renner's hypothesis about UNR positively affecting BACE1 translation is correct, UNR would still be available to generate the plaque precursor protein.

Interestingly, *APP* translation is driven by hnRNP C, like *UNR*, although it wasn't apparent whether or not this event involved the IRES in the *APP* transcript (Lee et al., 2010).

1.6 UNR and cancer

Between their seminal review of 2000 and subsequent update of 2011, Hanahan and Weinberg laid out a number of hallmarks of cancer (Hanahan & Weinberg, 2000, 2011). These included their six original hallmarks:

- Self-sufficiency in growth signals
- Insensitivity to antigrowth signals
- Limitless replicative potential
- Sustained angiogenesis
- Evading apoptosis
- Tissue invasion and metastasis

In terms of these hallmarks, UNR could immediately be linked to cancer in a general way. As an example, it could be postulated that a protein that controls the expression of a proto-oncogene like FOS, if mutated, could result in self-sufficiency in growth signals. Likewise, as UNR modulates the expression of the apoptosome scaffold protein, APAF1, a loss in that function could result in a cell potentially evading apoptosis. There are many such theoretical considerations that could link UNR to cancer but, in most cases, they are hypotheses based upon no experimental evidence whatsoever. That said, there is some circumstantial evidence in the literature that links UNR to cancer. For example, UNR has been shown to be differentially regulated or mutated in a number of cancers or during the development of the cancer phenotype.

Another of Hanahan and Weinberg's classic hallmarks is invasion and metastasis. The *UNR* transcript was shown to be down-regulated in cases of metastatic gastric cancer (p value = 0.00228 by random permutation test) compared to non-metastatic tumours. Biopsies of resected advanced gastric tumours were taken from 56 patients who were then followed up for three years. The patients were noted to have shown peritoneal metastases or to be recurrence free (as tested by peritoneal lavage). They were then assigned to one of two groups with 30 (13 with metastasis, 17 without) or 24 (13 with metastasis, 11 without) patients, respectively. The first group was designated the 'learning group' and microarray analysis of data from those patients was used to make predictions of which of 2304 genes could potentially be used as diagnostic markers of metastatic potential. These predictions were then tested by examination of the second 'validation group' samples (Motoori et al., 2006). Conversely, with advanced

cervical cancer treated with radiotherapy, it was found that *UNR* was expressed approximately two times higher in samples from patients who went on to die from cancer within 3-23 months compared to those who showed no signs of cancer for at least 39 months, going up to over 102 months, depending on the individual patients within the course of the study (Harima et al., 2004). The genomic region containing the *UNR* gene was found to be a common region to show a gain of copy number in paediatric adrenocortical tumours, the authors considering *UNR* to be important in adenomas but less so in carcinomas (Mateo et al., 2011). Interestingly, the region showed more frequent increases in copy number in carcinoma samples although the consensus length of amplification was slightly longer in carcinoma samples such that it included the *NRAS* gene.

1.6.1 UNR and cancer – COSMIC database search

There are other examples in the literature of UNR being associated with various cancers, mainly coming from high throughput experiments. The Catalogue Of Somatic Mutations In Cancer (COSMIC) database (version 70) listed 177 mutations in *UNR* found in different cancers, this equated to 95 changes in coding (Ray et al., 2015).

A more up to date version (v76) that used the GRCh38 version of the human genome was accessed on 16/02/2016 in order to obtain information about the amino acid mutations in UNR in cancer. It is worth pointing out that some observations from a given sample may have been coincidental and not related to the cancer, *per se*.

The first observation was that there appeared to be more arginines mutated than any other amino acid (Figure 1.6).

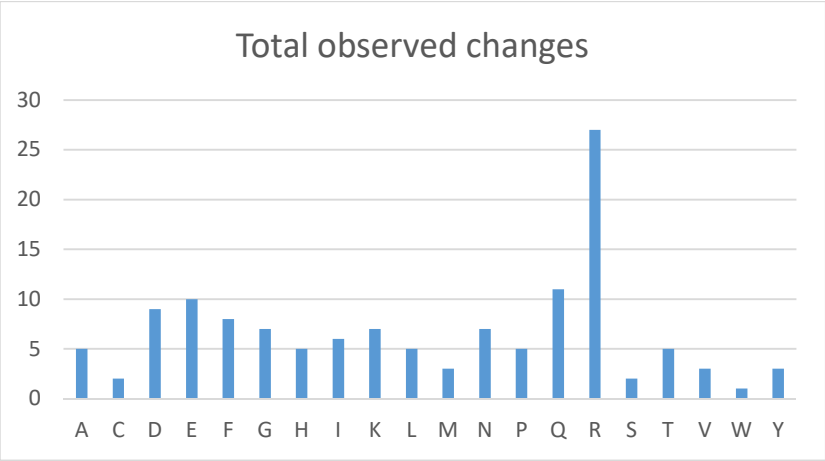


Figure 1.6: A bar chart showing the number of observed amino acid changes within the UNR protein recorded in the COSMIC database by affected amino acid residue. Changes include each mutation where the same residue has been recorded as being mutated into multiple amino acids and the introduction of stop codons but exclude silent mutations.

Taking the 844 amino acid version, there are 44 arginines in UNR, of which 20 were recorded as having at least one non-silent mutation. 17.1% of all recorded non-silent mutations were in arginine residues, more than both of the next most mutated amino acids (glutamine and glutamate at 7.7% each). This skew towards UNR having mutated arginines in cancer has been alluded to in the literature – e.g. it was noted that there are three recurrently mutated arginines within endometrial cancers (Bell, 2014). The reason for this overrepresentation of arginines is not known but it is easy to speculate that positively charged residues being mutated could alter, reduce or remove the capacity for UNR to bind to negatively charged molecules like RNA. Whilst pure speculation, it is considered unlikely that UNR-RNA binding would be rendered impossible as a result of these mutations. That is because it is believed that UNR binds RNA via its cold shock domains and these have particular highly conserved regions for binding nucleic acids (the RNP-1 [Y/FGFI] and RNP-2 [FFH] motifs) and neither of these contain arginine. It is postulated here, therefore, that the mutations in arginine residues are more about altering the thermodynamic stability of specific UNR-RNA interactions and, thereby, possibly altering the set of RNAs involved in UNR-RNA interactions.

Mutations recorded in the COSMIC database from within the five canonical cold shock domains of UNR are shown below (Figure 1.7).

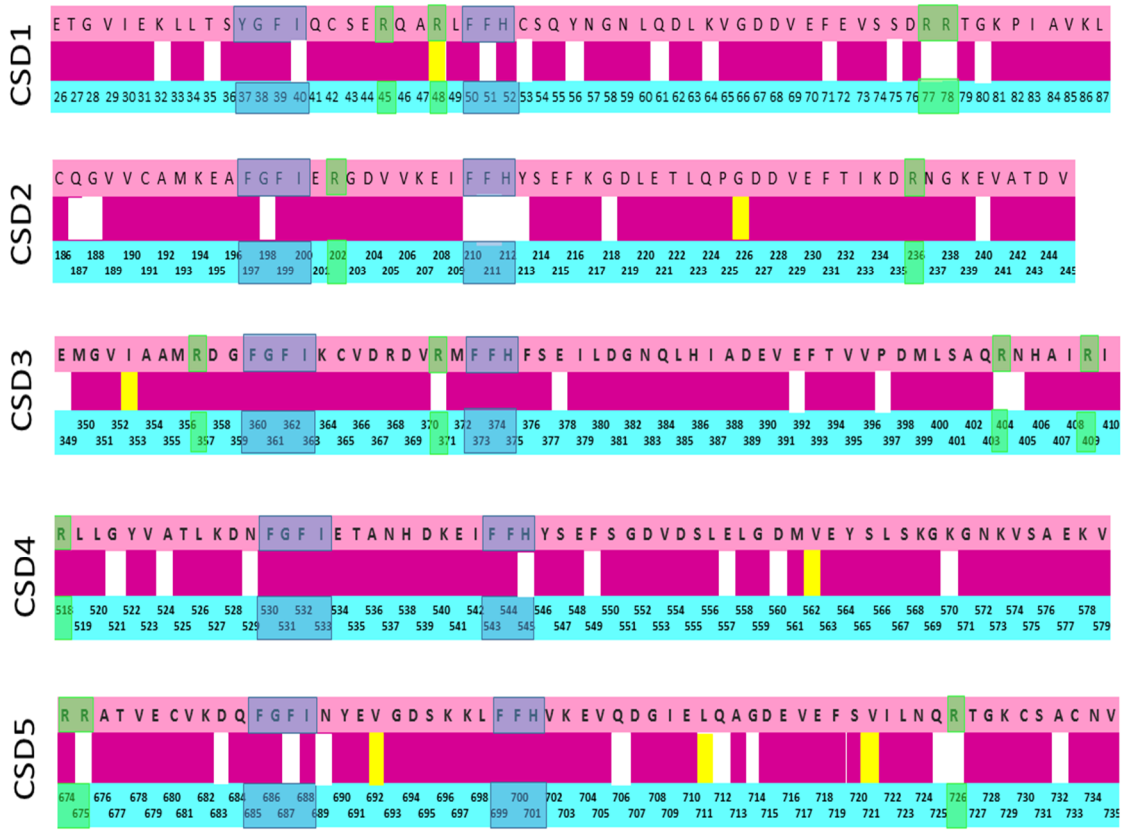


Figure 1.7: Schematic representation of the five canonical cold shock domains (CSD), as stated, within the UNR protein. Pink bands contain single letter amino acid code for the residues within each cold shock domain. Turquoise bands contain the numbers of the residues within the 798 residue protein. The magenta band refers to the individual residues. White spaces within the magenta bands refer to residues reported to have at least one non-silent mutation recorded in the COSMIC data referred to in the text. Yellow blocks refer to residues that have at least one silent mutation but no non-silent mutations recorded in the COSMIC data referred to in the text. Blue boxes highlight the two 'ribonucleoprotein domain' RNA-binding motifs (RNP1/2) found in each cold shock domain. Green boxes highlight arginines found within the cold shock domains.

Different cold shock domains appear to have different degrees of recorded mutations, an observation that holds for both the RNA binding motifs and the domains more generally (Figure 1.7). Cold shock domain 1 has recorded mutations in both of its RNA binding motifs. One is a conservative mutation I40V, whereas the other removes a conserved phenylalanine F51L. The possibility of arginine mutations towards the end of the domain is also observed. Cold shock domain 2 has some potentially catastrophic mutations in its RNA binding motifs. G198D takes away a small neutral residue and replaces it with a larger negatively charged residue that could

be expected to exert electrostatic repulsion on RNA molecules. Mutations have been observed in all residues of its RNP-2 motif (F210L, F211Y, H212D/H212Y). This could imply that alteration of the binding specificity of this CSD for RNA species is an important driver for cancers. It is worth considering the finding that CSD2 was important in the UNR-PABP interaction, although the mutated phenylalanine used in that work (equivalent to F199 here) did not have a mutation recorded in the COSMIC database (Ray & Anderson, 2016). The residues within the RNA binding motifs of cold shock domain 3 had no recorded mutations in the COSMIC database. It should also be considered here, however, that charged residues either side of RNP-2 have been reported to have been mutated which may affect the binding specificity of the domain for RNA species. Furthermore, as more cancer samples are sequenced and uploaded to the COSMIC database, new mutations in the RNA binding motif residues may be discovered. N529H, immediately preceding the residues in RNP-1, was recorded as a possible mutation in cold shock domain 4. The last conserved residue in RNP-2 (H545) has been recorded as being mutated to a tyrosine. A phenylalanine located just to the C terminal end of RNP-2 was recorded more than once as being mutated to a leucine (F549L). Cold shock domain 5 has a conservative mutation reported just N terminal of RNP-1 (D683E), as well as a mutation in RNP-1 itself (F687L). An N689S mutation has been found, abutting RNP-1 to the N terminal side. No mutations have currently been reported within RNP-2.

Another known region of UNR is its UNRIP binding site (v.i.). Looking for mutations within that site shows that some conserved UNRIP-binding residues are flagged up as potential sites of non-synonymous mutation in different cancers (Figure 1.8). The effect of these mutations on the UNR-UNRIP interaction, as well as any possible pathological changes in biological function, remain unknown. Given the huge array of functions carried out by UNRIP within the cell (see section 1.7), the potential effect of these mutations merits further consideration in future by other researchers.

Figure 1.8: Schematic representation of the reported Unrip-binding site in the C terminal region of UNR. The pink band contain single letter amino acid codes and the turquoise band contain the residue numbers within the 798 residue protein. The magenta band refers to the individual residues. White spaces within the magenta band refer to residues that have at least one non-silent mutation recorded in the COSMIC data referred to in the text.

It is unknown if any of the recorded mutations in the Unrip-binding site would prevent the binding of Unrip to UNR.

1.6.2 UNR and cancer – CONAN database search

The COpy Number ANalysis (CONAN) database (accessible from the COSMIC website) shows copy number changes recorded in cancer samples. It flagged up a number of copy number changes in *UNR*. An alignment of these data with similar data for *NRAS* showed that the *UNR* data was a subset of the *NRAS* data (data not shown). As *NRAS* is a known protooncogene and *UNR* is not, it was considered likely that the changes in copy number for *NRAS* would outweigh and confound any equivalent changes in *UNR* copy number. For this reason it was assumed to be unworthy of further consideration.

In another investigation, the expression profile of 4 pancreatic cancer cell lines and two pancreatic cancer primary cell lines were taken and compared with 11 libraries of non-neoplastic tissue that included both primary and cultured cells. The comparison showed 15 *UNR* tags per million in non-neoplastic samples and 144 *UNR* tags per million in the pancreatic cancer samples (Hustinx et al., 2004). Confoundingly, the tag they used for *UNR* was shared with that for lumican, a collagen binding protein reported to be upregulated in breast and colorectal cancer (Lu et al., 2002). As the same sequences were present in each transcript, obviously, they shared the statistics and it is impossible to infer anything about the proportion of relative upregulation accounted for by each transcript.

1.7 UNRIP

UNRIP is a known UNR-interacting protein (Hunt et al., 1999). The *Unrip* gene is believed to be expressed in all mouse tissues with high levels of expression observed in the liver and testicular tissue (Datta, Chytil, Gorska, & Moses, 1998). In comparison, *Unr* transcript levels were reported to be 7 times lower in liver compared to testis (Boussadia et al., 1997). The *UNRIP* gene is universally conserved among the eukaryotae from yeast to mammals (Datta et al., 1998), is often upregulated in cancer (e.g. in a variety of lung cancers, as compared with healthy lung tissue from the same patients (Halder et al., 2006)), and has been shown to be essential in mice (Chen,

Delrow, Corrin, Frazier, & Soriano, 2004). It is reported to be one of ten oncogenes responsible for the phenotype of one of the cancer cell lines used in this study – the SaOS-2 osteosarcoma cell line (Niforou et al. 2006).

Whilst UNRIP is not believed to possess a nuclear localization signal, it is observed in the nucleus (Halder et al., 2006). One protein with which it can interact is GEMIN7, part of the survival motor neuron (SMN) complex (Grimmler et al., 2005). The SMN complex is involved in splicing and consists of a variety of proteins including the eponymously named SMN protein and GEMINs 2-7 (Gubitz et al., 2004). Unrip was later shown to interact with the SMN complex via GEMIN7, although this relationship was only apparent in the cytoplasm. Knocking down *UNRIP* increased the nuclear localisation of the SMN complex (Grimmler et al., 2005). Interestingly, the interaction between Unrip and either GEMIN7 or UNR was noted to be mutually exclusive (Grimmler et al., 2005). Later work showed that the binding of another Unrip binding partner, La-related protein 6 (Larp6), was also mutually exclusive to both Gemin7 and Unr (Vukmirovic, Manojlovic, & Stefanovic, 2013). Protein sequences for UNR (O75534), LARP6 (Q9BRS8) and GEMIN7 (Q9H840) are freely available at www.uniprot.org. Manual comparison of these sequences suggests a shared consensus Unrip binding motif VLRxPRGPD (Figure 1.9). Importantly, a mutually exclusive relationship between UNRIP and either UNR or one of two other proteins implies that UNR function through UNRIP could be competitively reduced by the presence of LARP6 or GEMIN7. Also, UNR could competitively reduce the activity of UNRIP with LARP6 and GEMIN7.

Unrip interactor																									
Unr, C-terminal region	P	R	L	M	V	L	R	Q	P	R	G	P	D	N		S	M	G	F	G	A			E	R
Larp 6, C-terminal region	P		V	G	V	L	R	L	P	R	G	P	D	N		T	R	G	F	H	G	H		E	R
Gemin 7, N-terminal region	P		V	P	V	L	R	L	P	R	G	P	D	G	F	S	R	G	F	A	P	D	G		R

Figure 1.9: Conserved residues in mutually exclusive Unrip binding proteins imply a putative Unrip binding consensus sequence. Alignments were carried out by hand based on the www.uniprot.org human protein sequences O75534 (UNR), Q9BRS8 (LARP6) and Q9H840 (GEMIN7). Yellow shading signifies residue shared by all three proteins. Green shading signifies a serine or a threonine. Black shading signifies that a gap was intentionally inserted into a protein sequence to improve the alignment (subjectively) between the three proteins. N.B. This is similar to a diagram in Vukmirovic et al. (2013) although the amino acid range considered here is greater (their diagram runs from VRL to GF). They align the phenylalanine of GEMIN7 that lies immediately before the green-shaded serine with asparagines from the other two proteins – i.e. they shift the spaces from N-S/T to D-N).

UNRIP was shown to interact with Transforming Growth Factor beta (TGFβ) Receptor 1 and 2 (TGFβR1/2) homodimers in the presence or absence of receptor agonist (Datta & Moses, 2000). It was shown to interact with a variety of Sister of Mothers Against Decapentaplegic (SMAD) proteins; including SMAD2, SMAD3 and SMAD7 (Datta & Moses, 2000). UNRIP binding to SMAD7 increases SMAD7 repression of TGFβ signalling (Datta & Moses, 2000). One protein downregulated by UNRIP activity at TGFβ receptors is $p21^{Cip}$ and knocking down UNRIP increases $p21^{Cip}$ expression (Halder et al., 2006). Over-expression of UNRIP in cancer therefore leads to the hyperphosphorylation of the retinoblastoma protein, reducing its capacity to sequester proliferation-promoting E2F transcription factors, and driving cellular proliferation (Harbour & Dean, 2000).

The 3-phosphoinositide dependent kinase PDK1 binds to protein kinase B (PKB/AKT) and recruits it to the membrane where it can then act as one of the two PKB activating kinases. Other key targets include the ribosomal protein S6 kinase 1 (S6K1) and the atypical protein kinase C zeta (PKCζ) (Berridge, 2014). UNRIP/Unrip interacts with PDK1/Pdk1 and the interaction is enhanced by insulin signalling. The insulin effect is likely to be PI-3,4,5-P_3-dependent as it was not observed when cells were treated with wortmannin (Seong, Jung, Choi, Kim, & Ha, 2005). This indirect link between insulin signalling and UNR via UNRIP is worth considering in light of work linking UNR expression to control of diabetes. That work showed by qPCR that the *UNR* transcript level went up over 6 fold after one week of treatment in patients who were admitted to hospital with poorly controlled diabetes (Xavier et al., 2014). Interestingly, in light of current findings in this work linking UNR to selenium compound metabolism (see section 4.11.6), there were two other genes that were considered significant by the authors; one of which being Selenoprotein S, which fell more than three-fold following treatment (Xavier et al., 2014).

UNRIP has been shown to interact with the pro-apoptotic apoptosis signal regulating kinase 1 (ASK1), an interaction promoted by ASK1 phosphorylating UNRIP at T175 and S179. The addition of UNRIP was shown to decrease the amount of hydrogen peroxide-instigated apoptosis, whereas knocking down *UNRIP* increased the proportion of cells undergoing apoptosis in response to hydrogen peroxide (Jung, Seong, Manoharan, & Ha, 2010).

Drawing together the previous two sections, the pro-survival PDK1 was shown to interact with the pro-apoptotic ASK1 and that mutual phosphorylation reduced their respective downstream signalling activities (Seong, Jung, Ichijo, & Ha, 2010).

Unrip interacts with the Ewing's Sarcoma protein (EWS) in the nucleus, inhibiting an alternative interaction with the transcriptional co-activator EP300/CBP and downregulating EWS target genes such as *c-fos* (Anumanthan, Halder, Friedman, & Datta, 2006). As stated previously, UNR interacts with other proteins on the mCRD of the c-fos transcript, both stabilising and promoting degradation of the message following translation. EP300/CBP is required for EWS to activate hepatocyte nuclear factor 4 (HNF4)-mediated transcription (Araya et al., 2003). HNF4-mediated transcription is important in differentiation. For example, dedifferentiated hepatoma cells can be forced to re-differentiate following the stable expression of Hnf4 (Späth & Weiss, 1998). Gata6 is involved in transcribing the *Hnf4* gene within primitive endoderm and this is required for differentiation into visceral extraembryonic endoderm (Morrisey et al., 1998). As discussed previously, the essential function of Unr in mice was linked to destabilisation of the *Gata6* transcript (Elatmani et al., 2011).

UNRIP has been shown to interact with nuclear export factors and has also been suggested to be part of neuronal transport granules that transfer mRNAs into the cytoplasm and around the cell linked to microtubules (Tretyakova et al., 2005).

There are many other reported roles for UNRIP in the literature that shall be overlooked here for the sake of brevity. Based upon the lines of investigation that have been considered here, however, it appears that the functional roles of UNR and UNRIP overlap beyond any physical interactions between the two.

1.8 Concept of 'RNA operons'

When taken together, the current state of the literature pertaining to UNR implies that it is involved in controlling the cellular concentrations of a variety of mammalian proteins at the level of translation. The mRNA binding partners outlined here have included transcripts involved in universal processes like mitosis (e.g. *CDK11*), translation (e.g. *Unr* itself) and cellular proliferation (e.g. *Fos*). They also include mRNAs that encode proteins involved in essential metazoa-specific functions such as apoptosis (e.g. *APAF1*) and development (e.g. *Gata6*). Unfortunately, the literature concerning the role of UNR in the control of translation in human cells in incomplete. The current project aims to help address that gap by obtaining high-throughput data using a large-scale systems-based investigation.

Multiple related transcripts regulated by specific proteins have been termed 'RNA operons' (reviewed in Keene, 2007). The idea that UNR may be involved in modulating groups of transcripts that could be termed an RNA operon will be considered by GO-term analysis and other methods.

1.9 Use of sodium arsenite as a stressor

Sodium arsenite is used as a stressor in subsequent chapters. Arsenite has also been shown to modulate various signalling pathways at different concentrations across multiple cell lines (discussed in Bode & Dong 2002). Arsenite can downregulate EIF4E and, thereby, reduce cap-dependent translation (Othumpangat, Kashon, & Joseph, 2005).

Different effects of arsenite have been reported at different concentrations of the drug. Doses up to 1 µM were shown to be protective against subsequent oxidative stress and DNA damage in cultured skin cells (Snow, Sykora, Durham, & Klein, 2005). As little as 2-5 µM of arsenite were sufficient to lead to significant production of reactive oxygen species after 2 hours and significant depolarisation of the mitochondrial membrane after 30 minutes in MCF7 cells (Ruiz-Ramos, Lopez-Carrillo, Rios-Perez, De Vizcaya-Ruíz, & Cebrian, 2009).

0.1-1 mM arsenite was shown to reduce the rate of activity of the succinoxidase complex (Potter & Dubois, 1943), thereby showing it to be an inhibitor of cellular respiration. Later research showed that arsenite can generally inhibit the tricarboxylic cycle, with 50 µM arsenite able to reduce citrate production from pyruvate by 57% over control and 100 µM arsenite able to reduce citrate production from pyruvate by 100% in a rat cardiac mitochondria preparation (Reiss & Hellerman, 1958). Whilst not directly relevant to the cell lines used here, arsenite was also shown to also inhibit β-oxidation and ketogenesis in rat liver mitochondria (Rein, Borrebaek, & Bremer, 1979), thereby inhibiting pathways of energy production distinct from glycolysis->pyruvate->tricarboxylic acid cycle.

Several factors were taken into consideration when deciding on the concentration of the drug that was to be used. Experiments to calculate the LC50 of arsenite showed that 50% of HeLa cells died in 24 hours after treatment with around 200 µM arsenite, the lowest LC50 of the 4 cell

lines explored (Othumpangat et al., 2005). Other experiments had shown that UNR migrates to stress granules in HeLa cells following stress with 500 µM arsenite for 30 minutes (White & Lloyd, 2011) but that over 50% of keratinocytes die within 2 hours of exposure to 1 mM arsenite. Partially due to the findings of White and Lloyd (2011), and partially due to a desire to explore oxidative stress more than starvation, it was decided to use 500 µM arsenite for 30 minutes as a starting point in the experiments shown in this thesis. When immunofluorescence microscopy showed clearer staining for UNR in stress granules with minimal loss of cells when 1 mM arsenite was used for 1 hour, it was decided to use these conditions for the main experiments.

1.10 Aims and objectives

A brief glance over the literature pertaining to UNR shows a protein that is implicated in multiple biological processes of great interest to mankind such as mammalian development. Beyond academic interest, further research into UNR is potentially of economic benefit when it is viewed in terms of potential links to various pathological conditions that cause human suffering, decreased life expectancy and put strain on healthcare budgets around the world such as cancer, diabetes and Alzheimer's disease. In order to expand upon the current understanding of UNR function in human cells, this work intends to identify novel UNR-interacting proteins and mRNAs. It is hoped that the function of these currently unknown UNR-interactors may shed new light on the function of UNR in human cells.

2 Materials and methods

2.1 Materials

2.1.1 Antibodies

The antibodies used to generate Western blot images shown in this thesis* (Table 2.1), for carrying out immunoprecipitations/ribonucleoprotein immunoprecipitations (Table 2.2) and for carrying out immunofluorescence microscopy staining (Table 2.3) are presented below.

* For Figures 3.1 to 3.5 only, the antibodies used to detect UNRIP and/or UNR were the rabbit polyclonal antibodies documented in (Hunt et al. 1999), provided as a kind gift from Richard Jackson (University of Cambridge).

Table 2.1: Antibodies used for Western blotting

Target	Source	Dilution (v/v)	Company	Company ID	Storage
UNR	Rabbit pAb	1:1000	Novus Biologicals	NBP1-71914	4°C
UNR	Goat pAb	1:1000	Santa Cruz	sc-79292	4°C
UNRIP	Mouse mAb	1:1000	Santa Cruz	sc-136083	4°C
TP53	Mouse mAb	1:1000	Santa Cruz	sc-126	4°C
PABP	Rabbit pAb	1:1000	Abcam	ab21060	-20°C
SQSTM1	Goat pAb	1:1000	Santa Cruz	Sc-10117	4°C
HUWE1	Rabbit pAb	1:1000	Bethyl Laboratories	A300-486A	4°C
γ-tubulin	Mouse mAb	1:10000	Sigma	T6557	-20°C
α/β-tubulin	Rabbit pAb	1:2000	Cell Signaling	2148S	-20°C
Rabbit IgG	Donkey	1:10000	Santa Cruz	sc-2313	4°C
Goat IgG	Donkey	1:10000	Santa Cruz	sc-2056	4°C
Mouse IgG	Donkey	1:10000	Santa Cruz	sc-2318	4°C

N.B. Yellow shading denotes horseradish peroxidase (HRP)-conjugated secondary antibodies. The UNR IP that showed dirty staining for HUWE1 (Figure 4.19) was carried out using the antibody as stated in Table 2.2 with subsequent membrane probing with the anti-HUWE1 antibody listed in Table 2.1. There had previously been an anti-HUWE1 antibody raised in goat in the lab but that did not generate any bands when used to probe the membrane (see section 2.2.7). The membrane was then stripped (see section 2.2.8) and re-probed with the rabbit anti-HUWE1 antibody – thus the large amount of background staining.

Table 2.2: Antibodies used for immunoprecipitations and ribonucleoprotein immunoprecipitations

Target	Source	Amount used	Company	Company ID	Storage
UNR	Rabbit pAb*	5 or 6 µg per 25 µl Dynabeads**	Novus Biologicals	NBP1-71914	4°C
Control	Rabbit	5 or 6 µg per 25 µl Dynabeads**	Invitrogen	10500C	4°C

* The immunoprecipitation related to Figure 3.11 was carried out with Dr Swagat Ray using the anti-UNR immunoglobulin raised in goat (Table 2.1; see section 3.2.3).The associated conspecific IgG was obtained from Invitrogen (ID: 10200).

**IPs shown in chapter 3 used 5 µg per 25 µl but this ratio was calculated to be sub-saturating in terms of the amount of antibody being available to bind to the Protein A on the Dynabeads (according to the manufacturer of the beads [Invitrogen]), so 6 µg per 25 µl was used for main RIPs (see section 2.2.9 for further information).

Table 2.3: Antibodies used for immunofluorescence microscopy staining

Target	Source	Dilution (v/v)	Company	Company ID	Storage
UNR	Rabbit pAb	1:250	Novus Biologicals	NBP1-71914	4°C
UNR	Goat pAb	1:250	Santa Cruz	sc-79292	4°C
TIA1	Goat pAb	1:250	Santa Cruz	sc-1751	4°C
TP53	Mouse mAb	1:250	Santa Cruz	sc-126	4°C
HUWE1	Rabbit pAb	1:250	Bethyl Laboratories	A300-486A	4°C
Anti-Rabbit IgG	**Donkey**	**1:1000**	**Invitrogen**	**A21207**	**4°C**
Anti-Mouse IgG	**Donkey**	**1:1000**	**Invitrogen**	**A21202**	**4°C**
Anti-Rabbit IgG	**Donkey**	**1:1000**	**Santa Cruz**	**sc-362261**	**4°C**
Anti-Mouse IgG	**Donkey**	**1:1000**	**Santa Cruz**	**sc-362248**	**4°C**
Anti-Goat IgG	**Donkey**	**1:1000**	**Santa Cruz**	**sc-362245**	**4°C**

N.B. Coloured cells refers to the fluorophore conjugated to the highlighted secondary antibody (red = Alexa-594, light green = Alexa-488, dark green = CruzFluor™-488, blue = CruzFluor™-405)

2.1.2 siRNA

The siRNAs used in this thesis were:

1) si*UNR* (Life Technologies, ID = 122624, sequence from exon 8 of *UNR*)
2) control siRNA (Life Technologies, ID = Silencer negative control 2)

2.1.3 *PABP* primers used for reverse transcriptase-polymerase chain reaction amplification

The identity of these primers is currently unavailable to the author but should be available from Dr Emma Anderson upon her return to the University of Warwick.

2.1.4 Source of chemicals

Sodium arsenite was obtained from Sigma (S7400). All other laboratory chemicals were sourced from standard suppliers, including Sigma.

2.2 Methods

2.2.1 Tissue culture

HeLa, U2OS and SaOS-2 cells were all maintained in tissue culture dishes or flasks of various sizes depending on their ultimate use. They were fed Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% (v/v) foetal calf serum (FCS) (Biosera) and were incubated at 37°C (unless stated otherwise) in a humidified environment containing 5% (v/v) carbon dioxide in air.

Cells were split when they reached 70-100% confluent. To do this, the cells were washed twice with sterile phosphate buffered saline (PBS) at 37°C and treated with a trypsin/EDTA solution to

cause the cells to detach from their substratum and each other. They were then split 1:10 (HeLa) or 1:5 (U2OS and SaOS-2) into fresh medium.

2.2.2 Sodium arsenite treatment

Where cells were to be treated with sodium arsenite, they were fed fresh DMEM containing 10% (v/v) FCS that had been made with 100 mM sodium arsenite to a final concentration of 500 µM or 1 mM, as stated in the text. Negative control cells were mock treated with fresh DMEM containing 10% (v/v) FCS to which had been added the same volume of sterile PBS as 100 mM sodium arsenite had been added to the arsenite-treatment medium. The cells were then returned to the incubator for the amount of time stated in the text.

2.2.3 siRNA transfection

Both HeLa and U2OS cells were transfected with siRNA during the course of this thesis. They were transfected according to the same protocol which is laid out here using volumes appropriate for a single 10 cm plate.

1) Cells were split into 10 cm tissue culture plates and were allowed to grow to a confluency of 50-70%.
2) 1ml of Opti-MEM was mixed with 20 µl of Lipofectamine 2000 and left for 5 minutes at room temperature. Meanwhile, 2 µl of a 200 µM stock solution of either si*UNR* or control siRNA was added to another 1 ml of Opti-MEM.
3) After 5 minutes, the two solutions were mixed by inversion and left for 20 minutes at room temperature. This generated 2 ml of a solution of 200 nM siRNA.
4) During this time, the plates were fed with 18 ml of fresh DMEM containing 10% (v/v) FCS.
5) After the 20 minutes, the Opti-MEM mixture was added dropwise around the plates which were then returned to the incubator for 42-48 hours, giving a final siRNA concentration of 20 nM.

2.2.4 Immunofluorescence microscopy

Coverslips were prepared in a tissue culture hood by being dropped into 70% (v/v) ethanol (in sterile deionised water) before being removed and allowed to dry propped up on the side of a six well plate lid. Once dry, the coverslips were placed into wells of six well plates (one per well).

Cells were then split (see section 2.2.1) and diluted 1:10 (v/v) (HeLa) or 1:5 (v/v) (U2OS and SaOS-2) by surface area. Cells were then allowed to grow to a suitable confluency (generally 50-70%) before being processed for visualisation. Arsenite treatment was then undertaken if required. The cells were then washed twice with ice cold PBS and between 500 µl and 1 ml of ice cold 10% (w/v) methanal solution (made up as 10 ml of 37% (w/v) aqueous methanal diluted with 27 ml of phosphate buffered saline [PBS]) was then added to the wells to cover the coverslips. This was left for 10 minutes, after which the cells were washed twice with ice cold PBS. Between 500 µl and 1 ml of room temperature 0.5% (v/v) NP-40 solution (38 ml of PBS added to 2 ml of 10% (v/v) NP-40 in deionised water) was then added and left for 10 minutes, upon which time the coverslips were again washed twice with ice cold PBS.

300-500 µl of blocking reagent (serum from the species in which the secondary antibodies were raised – donkey [Sigma, D9663] for all experiments in this thesis – diluted to 2% (v/v) with PBS) was added directly to the coverslips and the plates were shaken gently for 60 minutes on an orbital shaker. The blocking reagent was then removed and 500 µl of primary antibody combinations, diluted in blocking reagent, were added to the appropriate coverslips without washing. The plates were again allowed to rotate gently on an orbital shaker for 60 minutes. The coverslips were then washed thrice with cold PBS and 300-500 µl of the required secondary antibody combination was added gently to each coverslip. The secondary antibodies were diluted in blocking reagent and were kept in the dark prior to application. The plates were then placed in tupperware boxes covered with aluminium foil and allowed to shake gently for 60 minutes. When 60 minutes had elapsed, the secondary antibodies were removed and the coverslips were washed thrice with cold PBS and placed cell side up against the edges of 6 well plate tops to dry under foil-covered tupperware for 15 minutes. During the 15 minutes, the necessary number of microscope slides were washed with 70% (v/v) ethanol in deionised water and allowed to dry.

When coverslips were dry, a drop of the Vectashield (Vector Laboratories), with or without DAPI as appropriate, was added to the centre of a microscope slide and a coverslip was placed on top, cell side down. The edges of the coverslip were then sealed with nail varnish and the slides were labelled and placed coverslip side up under foil-covered tupperware until the nail varnish set. When all slides were dry, they were placed in a slide holder, covered with aluminium foil and stored at 4°C prior to visualisation using a Leica SP5 confocal microscope with a 60X oil immersion objective. Images were then saved using the Leica confocal software associated with the microscope for off-site processing using LAS AF Lite software (Leica Microsystems).

In most cases, multiple optical slices were imaged and images are presented as max projections. Any changes in brightness or contrast were carried out equally on all channels in the images, which were saved as tagged image file format files. The images were then altered in the same way in Microsoft Office software (Powerpoint and Word) in order to enhance the visibility of the printed images.

2.2.5 Cell lysate generation

All lysates used in this thesis were generated using polysome lysis buffer (PLB). The recipe was taken from Keene et al. (2006). A pre-buffer was initially made and stored at 4°C prior to use. It consisted of:

100mM KCl,

5mM $MgCl_2$,

10mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) at pH = 7.0,

0.5% (v/v) NP-40.

Immediately prior to use, the following reagents were added to the stated end concentrations:

1mM dithiothreitol (DTT),

100 units ml^{-1} RNase OUT (Invitrogen),

400 µM vanadyl ribonucleoside complex (VRC),

1X protease inhibitor cocktail (Roche Diagnostics)

It should be noted that RNase OUT was used for the main experiments in chapters 4 and 5 but was not always used elsewhere (see text for further information).

Cells were washed twice in cold PBS before being just covered with cold PBS and gently scraped off the surface of the plates. The cell suspension was then centrifuged at 1000g and 4°C until the cells became sedimented at the bottom of the tube. The supernatant was removed and the cells resuspended in cold PBS and centrifuged as before and the supernatant was again discarded. Following a second resuspension and supernatant removal the cells were suspended in a similar volume of PLB and transferred to a suitably sized clean tube. This was incubated on ice for 5 minutes before being transferred to a freezer at -80°C for a period of no less than 8 hours. Finally, the raw lysates were thawed on ice before being spun down at 15000g and 4°C for 15 minutes, after which the supernatant was collected and the cell debris pellet discarded. The lysate was purified by a subsequent round of spinning down at 15000g and 4°C in order to separate out the aqueous portion of the lysate from unwanted lipids that were carried over from the raw lysate. A portion of the lysate was then set aside for protein quantification and the rest stored at -80°C until required.

2.2.6 Protein quantification

Protein quantification was carried out using Pierce BCA Protein Assay Kit (Thermo Scientific), as per the manufacturer's guidelines.

2.2.7 Western blot analysis

Cell lysates, if quantified, were diluted to the lowest protein concentration among them in pre-PLB buffer (see above) to give equal sample loadings. These lysates, or immunoprecipitated samples, were then heated to 95°C for 5 minutes in sodium dodecyl sulfate (SDS) loading buffer, which contained the following components at the following end concentrations:

50mM 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris) adjusted to pH=8.8 with concentrated hydrochloric acid.

10% (v/v) glycerol

2% (w/v) SDS

Bromophenol blue dye (used sparingly to generate a dark blue colour)

These samples were then subjected to polyacrylamide gel electrophoresis (Laemmli 1970). They were loaded onto an SDS polyacrylamide gel with a suitable protein marker and had their constituent proteins separated through a 5% stacking gel and a 10% resolving gel by electrophoresis. The gels were made using the reagents and quantities in Table 2.4.

Table 2.4 Reagents used in the formation of protein gels (by percentage of end mixture)

Component	Stacking	Resolving
Protogel (National Diagnostics)*	17	33.3
1M Tris (pH=6.8)	12.5	0
1M Tris (pH=8.8)	0	25
10% (w/v) ammonium persulfate	1	1
10% (w/v) SDS	1	1
TEMED	0.1	0.04
water	68	39.7

* 30% (v/v) acrylamide : 0.8% (v/v) methylene bisacrylamide

The gels were run at 120 V for 10 minutes followed by 200 V until adequate separation had been achieved. The proteins in the gel were then transferred onto nitrocellulose membrane using a wet transfer tank with a current of 350 mA for two hours in a room at 4°C or, alternatively, at a lower current overnight such that the end product of current and time was no less than 700 mA

hours. The membrane was then stained with Ponceau S dye (0.1% [w/v] of dye in 5% [v/v] ethanoic acid). The stained membrane was then photographed followed by being washed at least three times in TBST until the red colouration had gone. It was then blocked in 5% (w/v) reconstituted milk (Marvel, Premier Foods Group) in Tris-buffered saline-Tween20 (TBST) buffer with gentle rocking for 30 minutes. The milk was removed and a primary antibody (or, on occasion, two primary antibodies), diluted in 5% (w/v) Marvel in TBST, was then applied to cover the membrane which was then either left for 1-2 hours at room temperature with gentle rocking, or overnight at 4°C, again with gentle rocking. The primary antibody was then removed and the membrane was subjected to three 15 minute washes in TBST. A suitable secondary antibody, diluted in 5% (w/v) Marvel in TBST, was then applied to cover the membrane which was left to rock gently for 60 minutes. After this, the secondary antibody was discarded and the membrane was given five 15 minute washes in TBST with quick intervening swirl washes in TBST.

The membrane was then treated with an enhanced chemiluminescence (ECL) reagent made by the author based on a helpful suggestion by Dr Andrew Turnell (University of Birmingham, England) which was, in turn, based on (Haan & Behrmann 2007). The ECL solutions were made as follows:

Solution 1:

1) Make 10.5 ml 250mM luminol in dimethyl sulfoxide (DMSO)
2) Make 5 ml 90mM p-Coumaric acid in DMSO
3) Add 885.6 ml deionised water to 100 ml 1M Tris-HCl at pH = 8.5
4) Then add 10 ml of the luminol solution and 4.4 ml of the p-Coumaric acid solution
5) Wrap in aluminium foil, as luminol is light sensitive, label and store at 4°C.

Solution 2:

1) Add 899.36 ml deionised water to 100 ml 1M Tris-HCl at pH = 8.5
2) Add 640 µl 30% (v/v) hydrogen peroxide
3) Wrap in aluminium foil, label and store at 4°C.

Equal volumes of these solutions were mixed together immediately prior to applying to the membrane in a dark room and swirling for 60s. The ECL solution was then poured off and the membrane was placed protein side down on cling film and wrapped. It was then placed protein side up into a developing cassette. A photographic film was then placed on top of the membrane and the cassette was closed for a set amount of time. The film was then removed from the cassette and developed in an X-ray film processor. Depending on the appearance of the developed film, additional films were developed for longer or shorter periods of time. The films were then annotated, scanned and saved as .jpg files.

2.2.8 Stripping of membranes

Where a membrane was to be re-probed having been previously probed with a different primary antibody, Restore PLUS Western Blot Stripping Buffer (Thermo Scientific) was used, as per the manufacturer's guidelines.

2.2.9 Immunoprecipitation and ribonucleoprotein immunoprecipitation

PLB was used to generate all lysates which were used in immunoprecipitation experiments for which results are recorded in this thesis. As a result of this, the differences between standard immunoprecipitations and ribonucleoprotein immunoprecipitations were minimal (where stated, RNase OUT was not used in the lysis/immunoprecipitation buffer). The protocol was optimised in the Anderson lab by Dr Swagat Ray, based on Keene et al. (2006).

In brief, Protein A-conjugated Dynabeads (Life Technologies) were incubated with either an antibody against UNR or a control IgG (see section 2.1.1) in PBS containing 0.01% (v/v) Tween-20 (PBST) in tubes attached to a rotating mixer for 60 minutes at room temperature. Following this, the beads were blocked in 0.1% (w/v) bovine serum albumin (BSA) in PBS in tubes attached to a rotating mixer in a cold room at 4°C. Whilst the blocking step was carried out with every IP/RIP for which data is presented in this document, it was later discovered that the blocking step was, at best, unnecessary due to poor binding between the beads and BSA (publicity from the ThermoFisher website).

The beads were then washed twice in ice cold PBST and then once in ice cold NT2 buffer (50 mM Tris-HCl at pH=7.4, 150 mM NaCl, 1mM $MgCl_2$, 0.05% [v/v] NP-40). After this, they were resuspended in NT2 to which had been added: 5 µl of RNase OUT per ml, VRC to a final concentration of 400 µM and DTT to a final concentration of 1 mM. Cleared lysate was then added at 1 part to 9 parts bead suspension. The immunoprecipitation was then carried out in tubes attached to a rotating mixer for 4 hours at 4°C. The beads were then washed five times in ice cold NT2 using a magnetic stand to pellet the beads.

At this point, the beads were treated differently depending on the purpose of the IP experiment. Where the IP had been carried out for analysis by Western blot, the beads were boiled in SDS loading buffer. Likewise, where the IP was carried out to generate samples that were to be separated by SDS-PAGE prior to analysis by mass spectrometry, the beads were also boiled in SDS loading buffer (see section 2.2.10). Where the samples were to be used for the trial run using the on-bead method of sample preparation for mass spectrometry, all beads were immediately processed for that purpose (see section 2.2.11). Finally, in the case of the large scale experiments discussed in Chapters 4 and 5, the beads were separated 4 parts for RNA extraction and 1 part for analysis by mass spectrometry (on-bead sample preparation method).

2.2.10 Preparation of samples for mass spectrometry 1 – gel slice method

Immunoprecipitated samples were boiled in SDS loading buffer and their proteins were then separated by SDS-PAGE. The gel was then covered in InstantBlue (Expedeon) and left for 1 hour. Following this, the stained gel was processed for mass spectrometry. To do this, the lanes were cut into gel slices that were then transferred to labelled plastic tubes and these were then washed twice for 20 minutes in 50% (v/v) ethanol containing 50mM ammonium hydrogen carbonate (NH_4HCO_3) with shaking at 50°C. The supernatant was discarded after each wash. The gel was then dehydrated with 100% ethanol for five minutes at 55°C with shaking. Again, the supernatant was discarded. Disulphide bridges were then reduced with 10 mM DTT in 50 mM NH_4HCO_3 at 56°C for 45 minutes with shaking. The supernatant was discarded. Cysteine residues were then alkylated with 55 mM iodoacetamide in 50 mM NH_4HCO_3 for 30 minutes at room temperature in the dark. The supernatant was then discarded. Gel slices were then washed twice for 10 minutes each with 50% (v/v) ethanol in 50 mM NH_4HCO_3 with the wash liquid discarded on both occasions. The gel was then dehydrated with 100% ethanol for five minutes, after which the excess ethanol was discarded. Mass spectrometry grade trypsin, dissolved in

100 mM NH_4HCO_3, was then added to submerge the gel slices which were incubated overnight at 37°C with shaking. The same volume of 5% (v/v) methanoic acid in 25% (v/v) acetonitrile as that of the trypsin solution was then added and the samples were sonicated for 10 minutes in an ultrasonic cleaning bath with sweep mode (model number USC100T, VWR, Radnor, USA) prior to the supernatant being retained in a fresh labelled tube. This step was repeated twice with all supernatants for a given gel slice being merged. The peptide solutions were then dried overnight in a miVac DUO concentrator (Genevac Ltd., Ipswich, England) at 40°C. The pellets were then resuspended in 50 µl of 2.5% (v/v) acetonitrile containing 0.05% (v/v) 2,2,2-trifluoroethanoic acid (TFA) and the solutions were sonicated for 30 minutes. The samples were then centrifuged at 17000g for five minutes and the samples were transferred to labelled glass vials for further processing by the mass spectrometry department.

2.2.11 Preparation of samples for mass spectrometry 2 – on-bead method

It was possible for NP-40-free NT2-washed beads to be processed directly for analysis by mass spectrometry using the on-bead method.

In this, beads were re-suspended in 45 µl of 100mM NH_4HCO_3 and 5 µl of 100mM DTT was added. This mixture was incubated at 60°C for 15 minutes with shaking. 5 µl of 200 mM iodoacetamide was then added and the mixture was left in the dark at room temperature for 30 minutes. Mass spectrometry grade trypsin, dissolved in 100 mM NH_4HCO_3, was then added to cover and the samples were shaken overnight at 37°C. They were then centrifuged at 5000g for two minutes followed by the supernatants being transferred to Costar Spin-X columns (Corning) and centrifuged at 13300g for 15 minutes. The sample liquids were then transferred to glass vials for further processing by the mass spectrometry department.

2.2.12 Preparation of samples for mass spectrometry 3 – filter aided sample preparation

The knockdown experiments generated whole cell lysate samples in PLB as opposed to immunoprecipitated samples. These samples were processed directly for mass spectrometry as follows:

To around 80 µl of sample, DTT was added to an end concentration of 50 mM and the samples were heated with shaking at 55°C for 40 minutes. This was then added to 400 µl of buffer UA (8 M urea, 100 mM Tris-HCl at pH=8.8, made fresh) and mixed by vortexing. The solution was then centrifuged in 3k molecular weight cut off spin columns (Thermo Scientific) at 14000g for 40 minutes at room temperature. The eluent was discarded and a further 400 µl of UA was added to the column and it was respun as before. The eluent was discarded and 400 µl of iodoacetamide solution (46.2 mg iodoacetamide in 5 ml UA, end concentration 0.05 M) was added. This was mixed with shaking for one minute without heating and left to incubate for five minutes without shaking under aluminium foil. The column was then spun at 14000g for 40 minutes and the eluent was discarded.

The following step was then carried out three times: 400 µl of UB buffer (8 M urea, 100 mM Tris-HCl at pH=8.0, 10 ml ultrapure water, made fresh) was added to the columns which were centrifuged at 14000g for 40 minutes with the eluent discarded.

The following step was then carried out three times: 400 µl of 100 mM NH_4HCO_3 was added to the columns which were centrifuged at 14000g for 40 minutes with the eluent discarded.

Mass spectrometry grade trypsin, dissolved in 100 mM NH_4HCO_3, was then added to the columns which were shaken for one minute before being transferred to a 37°C water bath overnight. The columns were then transferred to new labelled tubes and spun down at 14000g for 40 minutes. The eluent was not discarded. 50 µl of 0.5 M NaCl was then added to the columns and then were spun down again at 14000g for 20 minutes. The columns were then discarded and TFA was added to the tubes to an end concentration of 0.5% (v/v). As discussed at the relevant section in the text, the protocol was changed by the mass spectrometry department during the course of this work due to an error in the concentration of acid added at this point. Previously, the protocol had stated that 1 µl of 5% (v/v) TFA should be added here. As there was 120 µl of trypsin solution and 50 µl of NaCl present, the end concentration would be less than 0.03% - more than ten times lower than was suggested in their updated protocol.

Rows of tubes were then arranged that had 200 µl of acetonitrile in the first row, the samples in the second row, 200 µl of 0.1% (v/v) TFA to wash and desalt in the third row and 50 µl of 0.1% (v/v) TFA in 50% (v/v) acetonitrile for elution in the fourth row. A C18 ZipTip (Millipore) was then

activated by pipetting up and down the acetonitrile 12 times. The activated tip was then inserted into the sample below and peptides were collected by pipetting up and down 12 times. The peptides were washed and desalted by pipetting the TFA in acetonitrile below up and down 12 times. Finally, the peptides were eluted by pipetting the elution buffer below up and down 12 times. This was repeated with a fresh tip for each sample.

The samples were then spun in the Speed-Vac for 150 minutes (-OH setting at 40°C) to evaporate all the solvent. The samples were then resuspended in 50 µl of ultrapure water, to which was added 500 nl of 10% (v/v) methanoic acid. They were finally spun down at 17000g for three minutes and transferred to labelled glass vials for further processing by the mass spectrometry department.

2.2.13 Extraction of RNA

RNA was extracted from whole cell lysate or immunoprecipitated material using Trizol reagent (Life Technologies) as follows:

1) 1 ml of Trizol added to beads or 100 µl of cell lysate
2) Vortex and leave at room temperature for five minutes
3) Add 200 µl of trichloromethane
4) Vortex for 15 seconds and leave at room temperature for two minutes
5) Centrifuge at 12000g for 15 minutes at 4°C
6) Transfer aqueous layer to clean tube and add 500 µl of trichloromethane
7) Vortex for 15 seconds and then centrifuge at 12000g for 15 minutes at 4°C
8) Transfer aqueous layer to clean tube
9) Add 1 µl of 20 mg/ml glycogen (Invitrogen, ID = 10814-010) and mix by finger tapping
10) Add 500 µl of propan-2-ol and incubate for 10 minutes at room temperature
11) Centrifuge at 12000g for 15 minutes at 4°C and then discard supernatant
12) Add 1 ml of 75% (v/v) aqueous ethanol and store for at least eight hours at -20°C
13) Dislodge pellet by inversion
14) Centrifuge at 7500g for five minutes at 4°C
15) Remove most of the supernatant and spin down at 7500g for five minutes at 4°C
16) Remove remaining supernatant with p20 pipette and leave pellet to dry under a hood
17) Re-suspend pellet in 15 µl of nuclease-free water by pipetting
18) Incubate at 50°C for ten minutes
19) Transfer to a clean tube and store at -80°C until required

2.2.14 Reverse transcriptase polymerase chain reaction amplification of cDNA

Reverse transcriptase polymerase chain reaction amplification of cDNA and subsequent visualisation was carried out as follows:

1) 10 pmol each of forward and reverse *PABP* primers were mixed with with nuclease-free water and Biomix Red (Bioline, to an end concentration of 1x).
2) To 18 µl of this mixture was added 2 µl of test solution (cDNA from a UNR IP, cDNA from an IgG IP, a total cell cDNA positive control or a nuclease-free water negative control).
3) The mixes were then subjected to amplification in a PCR machine with the following conditions:

    1) Heat lid to 105°C
    2) 30s at 94°C
    3) 30s at 57°C
    4) 30s at 72°C
    5) Repeat steps 2-4 a further 29 times
    6) Hold at 4°C until samples are removed

4) Amplified cDNA samples subjected to electrophoresis on an agarose gel (2% in Tris-Acetate-EDTA buffer [40 mM Tris, 20 mM ethanoic acid, and 1 mM EDTA at pH=8]) containing ethidium bromide with a DNA ladder. The gel was run at 100 V/60 mA for 90 minutes in the same buffer.

5) The gel was then imaged in a UV box attached to a printer.

The experiment presented in section 3.2.3 was carried out with assistance from Dr Swagat Ray using cDNA he had generated previously. He had made the cDNA from RNA he obtained from:

a) HeLa cell lysate and,
b) UNR/IgG IPs using the same lysate as in (a).

2.2.15 Preparation of samples for RNA-sequencing

The RNA obtained in section 2.2.13 was handed over to the Genomics Department within the School of Life Sciences, University of Warwick. It was then processed according to their standard operating procedures and passed on to a Wellcome Trust site in Oxford, England for sequencing. Sequencing was carried out on an Illumina HiSeq4000 on a 75 base pair paired end run according to their own standard operating procedures. The details of the processing carried out within Warwick is detailed briefly in section 5.1.1.

2.2.16 Analysis of mass spectrometry results using Scaffold

The proteomics department in the School of Life Sciences of the University of Warwick provided .sf3 Scaffold files with a semi-quantified summary of the data.

A free Scaffold file viewer offered for download from the Proteome Software website was used to visualise the data contained in the .sf3 files. Some figures contained in this work were generated by the viewer software. Decisions pertaining to statistical significance, in experiments that included repeats, were taken in two parts (hits by t test p value and hits by fold enrichment) as follows:

a) Hits based on t test p values

1) Scaffold-generated abundances were exported to Microsoft Excel
2) Two tailed paired t tests were carried out at the level of the individual protein
3) Those proteins with p values below a stated cut off were taken to be significant

b) Hits based on fold enrichment (immunoprecipitations)

1) A ratio was generated of the average values for the UNR and IgG samples by protein*
2) Those proteins with ratios >10 and <∞ were considered significant
3) Those with ratios =∞ were considered significant if all UNR abundances were >0

* UNR = immunoprecipitated samples using an anti-UNR immunoglobulin, IgG = negative control immunoprecipitated samples using a conspecific nonspecific IgG.

c) Hits based on fold enrichment (knockdowns)

1) A ratio was generated of the average values for the siU and siC samples by protein*
2) Those proteins with ratios >10 and <∞ were considered significant
3) Those with ratios =∞ were considered significant if all siU abundances were >0
4) Those proteins with ratios <0.1 and >0 were considered significant
5) Those with ratios =0 were considered significant if all siC abundances were >0

* siU = cell lysates formed from si*UNR*-treated cells, siC = cell lysates formed from control siRNA-treated cells

2.2.17 Analysis of mass spectrometry results using Progenesis

The proteomics department provided raw data files generated by their mass spectrometry machine in .raw format. Due to licencing restrictions, these were loaded into the 'Progenesis QI for Proteomics' software (Nonlinear Dynamics) within the mass spectrometry department. Following the identification of putative peptide sequences, these were exported to the MASCOT server (Matrix Science) for assignment to proteins. The MASCOT searches were then reimported

into Progenesis which then quantified proteins based on the peptides assigned to them by MASCOT.

As with Scaffold, some figures were exported directly from Progenesis for use in this thesis. The methodology for detecting significant proteins was also the same (abundances exported to Excel for analysis by t test, etc.).

2.2.18 Analysis of RNA sequencing data using DESeq2

The RNA sequencing data was pre-processed by Dr Nigel Dyer (University of Warwick, England) and presented to the author as a Microsoft Excel file of abundances, as described in section 5.1.1. MATLAB was then used to carry out principal component analysis one the samples to detect outliers and verify that the majority of the variation in the remaining samples was associated with known sources such as cell type or treatment. This is discussed in section 5.2. The samples that were not excluded as outliers (or as the si*UNR*-immunoprecipitation or control siRNA-immunoprecipitation pairs of outliers – see section 5.3.2) were analysed using the R tool DESeq2 (Love et al. 2014). DESeq2 was used to generate Benjamini-Hochberg-adjusted p values for RNAs that had been detected by RNA sequencing. The precise ways in which DESeq2 was used are described in more detail in the relevant chapter.

2.2.19 Gene Ontology term overrepresentation analysis

Gene Ontology (GO)-term overrepresentation analysis was carried out on significant proteins or transcripts, detected as in the previous sections, from different comparisons. The protein/transcript lists and/or the technical data for the analyses are provided at the relevant sections.

The gene ontology (GO) tool chosen to analyse proteins (Chapter 4) or protein coding RNAs (Chapter 5) that were pulled down in RIPs, or whole cell proteomes (Chapter 6) was the 'AmiGO 2 Term Enrichment Service'. This tool records the total number of genes associated with a given GO term in the reference genome (the background frequency). It predicts how many genes out

of a sample of a certain size should be associated with a given GO term in a purely random sample. It then notes the actual number of genes in the sample that do map to the given GO term (the sample frequency) and calculates the probability of observing at least that number of genes being observed by chance alone, assuming the sample was indeed randomly selected.

Associated gene names were fed into the tool and, where it failed to recognise certain genes, those genes were removed. The GO tool contained multiple entries for some inputted gene names that referred to different genes or proteins. Where that occurred, the alternative GO tool-suggested gene identifiers or Uniprot protein identifiers were compared against additional information provided by Scaffold, Progenesis or DESeq2. That allowed for a Uniprot identifier to be located that referred to a protein encoded by the gene in question. The ambiguous gene name was then replaced with the Uniprot identifier and the updated list was re-entered into the GO tool.

The Bonferroni multiple testing correction was used in all tests and these were carried out for each of the following three classes of GO term:

1) Biological Process
2) Molecular Function
3) Cellular Component

3 Validation of techniques for investigating UNR expression and localisation

In order to proceed with the high-throughput experiments that form the main body of this work, it was considered essential to demonstrate initially that the groundwork methods worked with respect to the available reagents and equipment. For example, failure to demonstrate that UNR could be immunoprecipitated from cell lysates with the available antibodies would call into question any mass spectrometry work to identify UNR-interacting proteins from UNR co-immunoprecipitated samples.

This initial experimental chapter therefore set out to confirm that UNR could be detected by Western blotting and immunofluorescence microscopy. It furthermore set out to confirm that UNR could be immunoprecipitated from HeLa whole cell lysate in combination with a known UNR-interacting protein and mRNA (PABP and *PABP*, respectively). Finally, the effect of sodium arsenite treatment on the distribution of UNR in three different cell lines was to be explored.

Re: section 3.2.3, the experiment was carried out with Dr Swagat Ray who made the original image shown in Fig 3.11 and had prepared the cell lysate used for the immunoprecipitation. Re: section 3.2.4, the experiment was carried out with Dr Swagat Ray using cDNA that he had prepared previously.

3.1 UNR and UNRIP can be detected by Western blotting

3.1.1 Western blot analysis can detect recombinant UNR and UNRIP

As the lab possessed some legacy recombinant UNR protein and legacy antibodies, it was considered useful to show that the antibodies could detect recombinant proteins. The recombinant proteins were:

1) UNR
2) UNR/UNRIP mixture

Samples were prepared by boiling 50 ng, 30 ng, 20 ng and 10 ng aliquots of each set in SDS loading buffer and running these on a 10% polyacrylamide gel prior to transferring the proteins to nitrocellulose membrane. The membrane was then blocked and probed with antibodies against UNR (rabbit polyclonal, see section 2.1.1), UNRIP (rabbit polyclonal, see section 2.1.1) or a combination of the two, as appropriate. Horseradish peroxidase (HRP)-conjugated anti-rabbit IgG secondary antibodies were then applied and the membrane was ultimately developed using an enhanced chemiluminescence-based approach.

This showed that the antibody was able to detect UNR but that the pure recombinant UNR protein had become degraded over time (Figure 3.1).

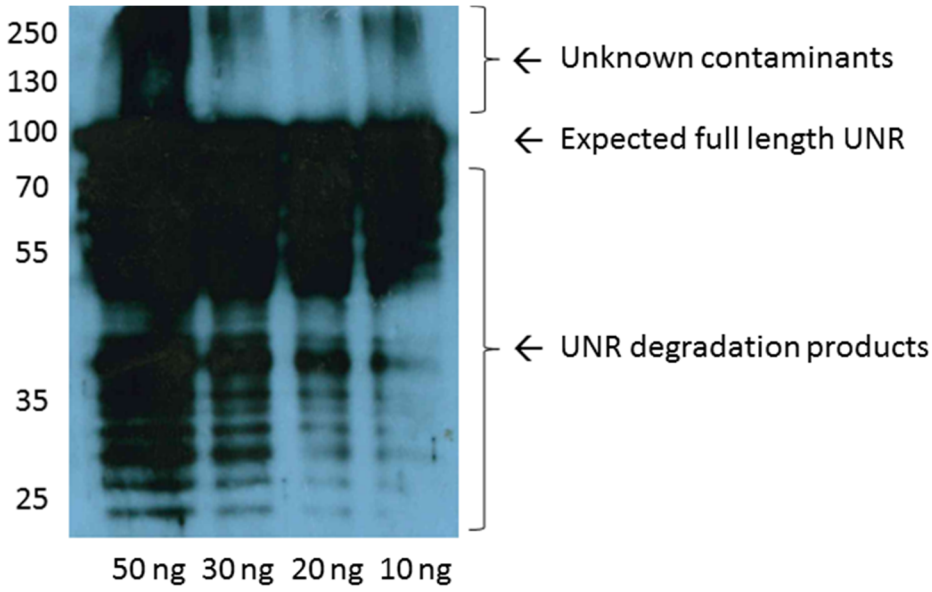


Figure 3.1: Result of a Western blot probing for recombinant UNR protein boiled in loading buffer. 10 ng, 20 ng, 30 ng or 50 ng (as stated) of the protein was loaded per lane of a 10% polyacrylamide gel prior to being transferred onto nitrocellulose membrane which was blocked, probed with a polyclonal anti-UNR immunoglobulin raised in rabbit, followed by a HRP-conjugated anti-rabbit IgG secondary antibody and then developed using an enhanced chemiluminescence-based approach. The numbers on the left hand side of the image are approximate molecular masses (in kilodaltons) as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher).

The recombinant UNRIP protein showed little degradation as part of the UNR/UNRIP mixture (Figure 3.2). Interestingly, whilst the recombinant UNR in the UNR/UNRIP mixture was degraded, it appeared to be much less so than the pure sample (Figure 3.2).

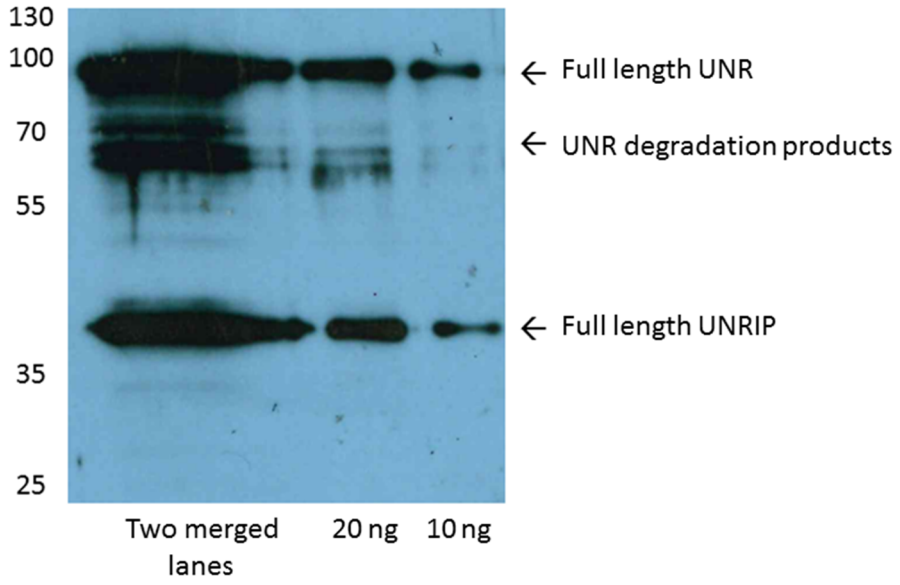


Figure 3.2: Result of a Western blot probing for UNR and UNRIP. A mixture of recombinant UNR and recombinant UNRIP proteins were boiled in loading buffer and the following amounts were added to wells of a 10% polyacrylamide gel: left to right 50 ng, 30 ng (these two lanes ran into each other), 20 ng and 10 ng (as stated). The mixtures were then separated by SDS-PAGE prior to being transferred onto a nitrocellulose membrane which was blocked, probed with a mixture of anti-UNR and anti-UNRIP immunoglobulins raised in rabbit (see section 2.1.1) followed by a HRP-conjugated anti-rabbit IgG secondary antibody and then developed using an enhanced chemiluminescence-based approach. The numbers on the left hand side of the image are approximate molecular masses (in kilodaltons) as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher).

3.1.2 Western blot analysis can detect endogenous UNR and UNRIP

Having demonstrated that UNR and UNRIP could be detected by Western blotting using recombinant proteins, it was decided to show that these proteins could also be detected by Western blotting from HeLa cell lysate.

50 µg of HeLa cell lysate that had been produced and quantified in the lab previously by Dr Swagat Ray was boiled in SDS loading buffer and run on a 10% polyacrylamide gel prior to the proteins being transferred to a nitrocellulose membrane. The membrane was then blocked and probed with antibodies against UNR and UNRIP (see section 2.1.1). A HRP-conjugated anti-rabbit IgG secondary antibody was then applied and the membrane was ultimately developed using an enhanced chemiluminescence-based approach (Figure 3.3).

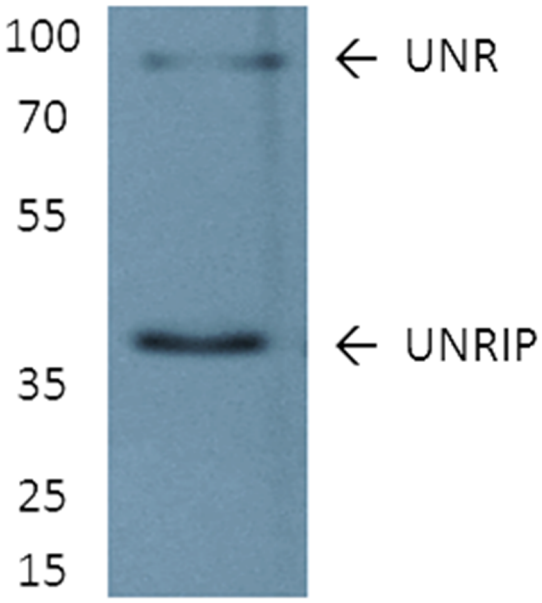


Figure 3.3: Result of a Western blot probing for endogenous UNR and UNRIP proteins using HeLa cell lysate. 50 µg of lysate was boiled in SDS loading buffer and loaded in a lane of a 10% polyacrylamide gel prior to being transferred onto nitrocellulose membrane which was blocked, probed with a mixture of anti-UNR and anti-UNRIP immunoglobulins raised in rabbit (see section 2.1.1) followed by a HRP-conjugated anti-rabbit IgG secondary antibody and then developed using an enhanced chemiluminescence-based approach. The numbers on the left hand side of the image are approximate molecular masses (in kilodaltons) as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher). This is the last figure that was made using the UNRIP antibody that was gifted to Emma Anderson by Richard Jackson (see section 2.1.1).

3.1.3 UNR levels decrease with increasing cell confluency in HeLa cells

It was stated in the literature that UNR levels vary in a cell cycle dependent manner and that this observation can result in UNR having cell cycle-dependent functions (Tinton et al. 2005; Schepens et al. 2007). In order to ascertain the strength of the cell cycle dependent expression of UNR effect, it was decided to carry out a confluency course experiment looking at the amount of UNR in lysates obtained from HeLa cells grown to different confluencies. It was assumed that plates with a lower confluency of cells would be expected to have a larger proportion of cells undergoing mitosis. This, in turn, could be expected to lead to a larger relative yield of UNR than harvesting the same number of cells from confluent plates where very few, if any, cells would be expected to be undergoing mitosis.

It had been decided not to synchronise cells prior to harvesting in order to obtain a general overview of UNR function that was, as much as possible, independent of cell cycle specific functions. That was with a view to carrying out later mass spectrometry and RIP-sequencing experiments with samples obtained from UNR immunoprecipitation. In order to confirm whether or not any apparent change in the level of UNR was due to unequal loading, a photograph was taken of the Ponceau S-stained membrane prior to developing the UNR images and the membrane was later stripped and re-probed for γ-tubulin.

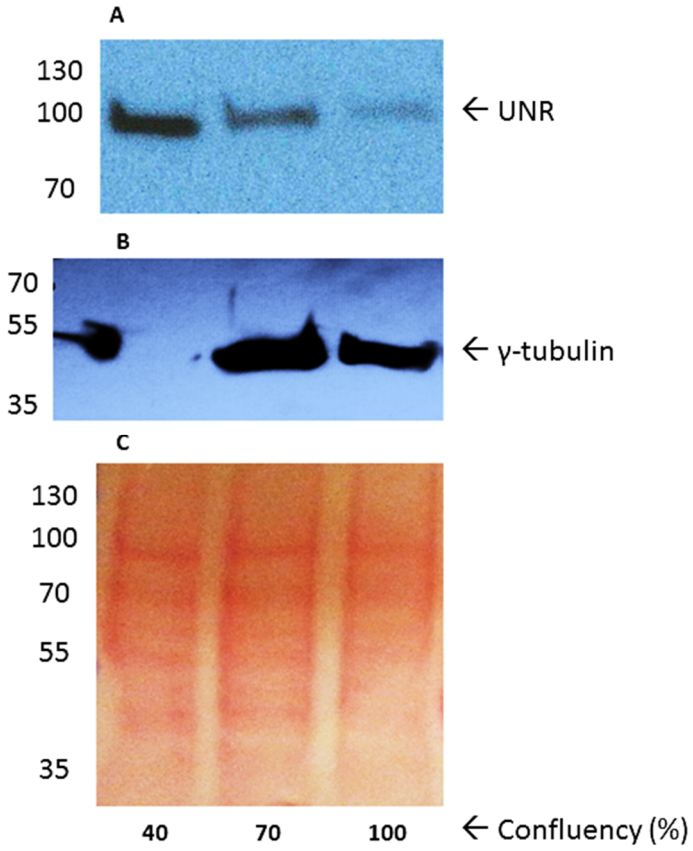


Figure 3.4: Western blot detection of UNR (A) or γ-tubulin (B) and Ponceau S staining of the original membrane (C), by confluency at the point of harvesting HeLa cells, as stated. 200 µg of total protein was loaded per lane. The membrane was originally probed for UNR, then stripped and re-probed for γ-tubulin. The numbers on the left hand side of the image are approximate molecular masses (in kilodaltons) as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher).

It seemed clear that the γ-tubulin blot was not a fair loading control as, if it were, it would have to be assumed that there was no protein in the first lane even though there was a strong UNR band and Ponceau S staining in that lane, and that there was γ-tubulin in the gel marker (Figure 3.4). Comparing the γ-tubulin image with the Ponceau S staining showed clear differences. The

Ponceau S staining showed that there was a similar amount of protein in all the HeLa lanes, as would be expected if protein quantification had been carried out correctly. A quick search of the literature suggested that proteins encoded by housekeeping genes can fluctuate in concentration with fluctuations in cell confluency (Greer et al. 2010). For this reason, it was decided to use Ponceau S staining as a loading control for many subsequent Western blots, especially those in which there could be differences in confluency between the samples (Romero-Calvo et al. 2010).

In terms of the relative amounts of UNR in cells at different confluencies, it appeared that there was more UNR in HeLa cells that were at lower confluency (Figure 3.4). This posed an issue for subsequent experiments in which large amounts of lysate was to be obtained. It could be that harvesting cells at lower confluency could generate more UNR even though fewer cells would contain less protein overall. Cell type specific differences in this pattern (compare Figures 3.4 and 3.15) and concerns over the risk of not getting a fair sample of all potential UNR-containing complexes, led to the decision to harvest cells when they were around 70% confluent. It was hoped that that would give a better spread of mitotic and non-mitotic cells and, hopefully, a more representative sample of UNR-containing complexes.

Another observation was that UNR levels appeared to increase in HeLa cells that were allowed to become over-confluent (data not shown). It was decided not to include images of this as the conditions were non-physiological and potentially difficult to replicate (i.e. localised cell density would change the amount of oxygen and nutrients available to cells attached to the surface of the plate meaning that lysate made from 110% confluent cells with one large clump could have a different composition to lysate made from 110% confluent cells with the extra 10% spread out more sparsely over the monolayer of cells attached to the plate).

3.1.4 Cold shock may reduce UNR levels in confluent HeLa cells

It was stated in section 1.3.1 that cold shock domains are associated with response to cold shock in bacteria but that there was no evidence for UNR being involved in response to cold shock in mammalian cells. It was decided to test the assumption that UNR is not involved in the cold shock response in human cells by exploring whether or not cold shock had an effect on the UNR protein level in confluent HeLa cells. To do this, four plates of HeLa cells were grown to near confluence in a 37°C humidified incubator with 5% carbon dioxide. Half were then fed with fresh

medium that had been heated to 33°C and were transferred to a 33°C humidified incubator with 5% carbon dioxide. The other two were fed with medium at 37°C and returned to the 37°C incubator. They had been allowed to reach near confluence as previous experience had shown that HeLa cells take a few days before becoming over-confluent. As the cells were expected to slow their growth at 33°C, this was meant to ensure that the cells would be at similar confluency at the end of the experiment and that there wouldn't be a major change in the percentage of cells that were in G_0. One plate from each incubator was harvested at 4 hours and the others at 9 hours (Figure 3.5).

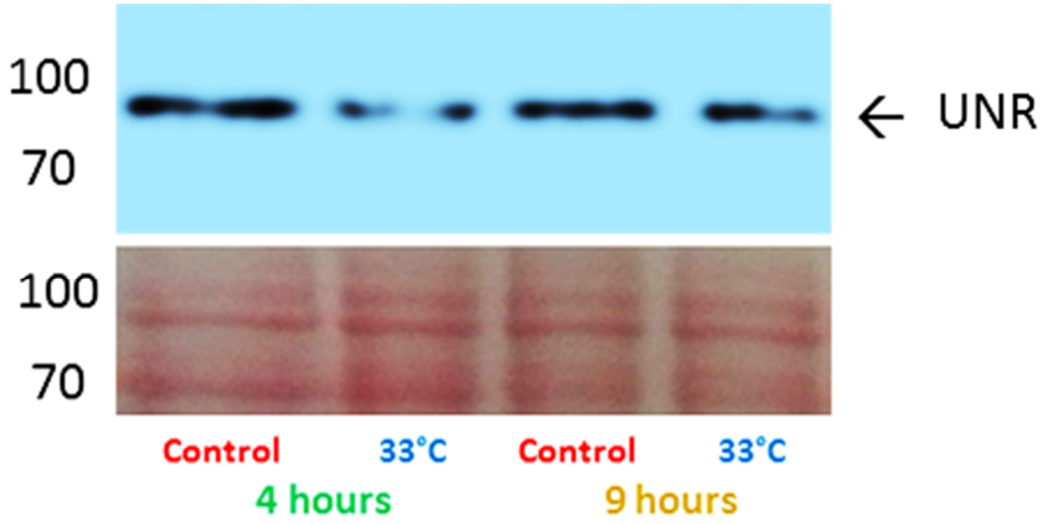


Figure 3.5: Western blot showing UNR. Four plates of almost confluent HeLas cells were grown at 37°C before being separated at time=0 and fed with fresh DMEM (containing 10% FCS) that had been heated to either 33°C or 37°C. The two plates fed with medium at 37°C were returned to the incubator set at 37°C and the other two were moved to a separate incubator at 33°C. Other conditions were kept the same. Cells from one plate from each set were harvested at the specified times, the lysates were purified and quantified and 50 µg total protein per lane was run on a 10% polyacrylamide gel. The upper panel shows the result of the Western blot; the lower panel shows a photograph of Ponceau S staining of the membrane over the same region as shown in the upper panel. The numbers on the left hand side of the image are approximate molecular masses (in kilodaltons) as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher). This is the last figure that was made using the UNR antibody that was gifted to Emma Anderson by Richard Jackson (see section 2.1.1).

In order to calculate the relative changes in the amount of UNR between the different conditions, Image J was used to measure the intensity of the UNR bands. Comparing the cold shock samples to the corresponding control samples, there was around 38.5% less UNR in the cold shock sample at 4 hours and around 24.1% less at 9 hours. This implies that cold shock may reduce UNR levels in confluent HeLa cells.

This experiment was repeated with cells split and grown at 33°C but they grew at a much slower rate and it was not considered fair to compare them with cells that were grown at 37°C for a different number of days or, alternatively, were harvested several days sooner or at a different confluency (data not shown).

3.1.5 Arsenite stress causes UNR but not TP53 to become localised in punctate structures within 30 minutes

Results in the literature and from the lab suggested that UNR is found in stress granules following arsenite stress (White & Lloyd 2011; Ray, Ó Catnaigh & Anderson [in preparation]). As will be mentioned later, results in the lab also linked UNR to TP53 (see below). It was therefore decided to carry out an immunofluorescence microscopy-based investigation into the distribution of UNR and TP53 in HeLa cells that were either unstressed or stressed with 500 µM sodium arsenite for 30 minutes. To do this, HeLa cells were grown on cover slips before treated with 0.5 mM sodium arsenite in fresh DMEM (containing 10% FCS) or mock treated with a similar volume of sterile PBS in fresh DMEM (containing 10% FCS) for 30 minutes. After this time, the cells were fixed, permeabilised, blocked, treated with primary and then secondary antibodies before being mounted with DAPI (Figure 3.7). Subsequent immunofluorescence imaging was carried out with or without DAPI as shown in each figure. Prior to carrying out analysis looking for UNR and other specific proteins, it was considered useful to carry out a control experiment using stressed and unstressed HeLa cells, treated as above but without primary antibodies. This was intended to give an idea of the background staining expected of our fluorophore-conjugated secondary antibodies (Figure 3.6).

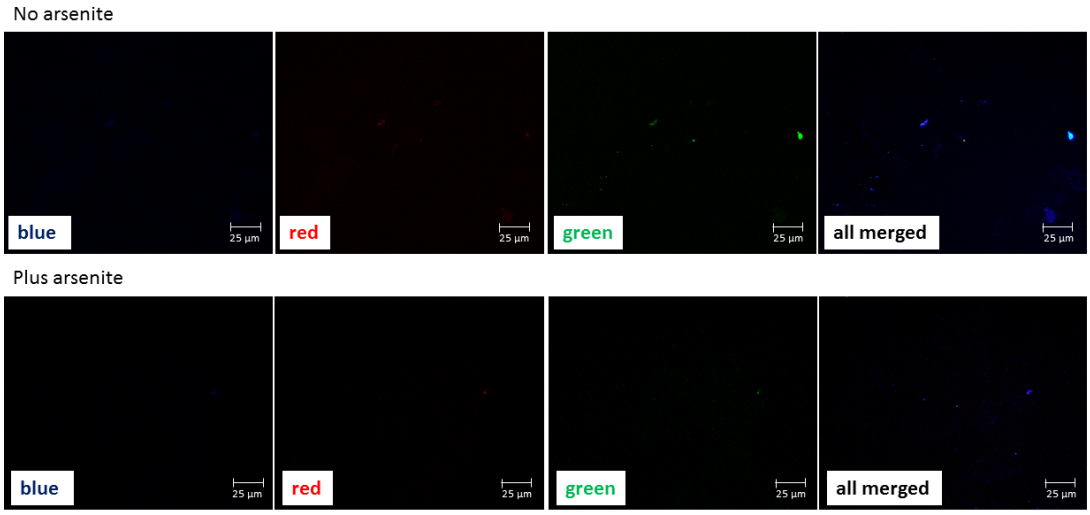
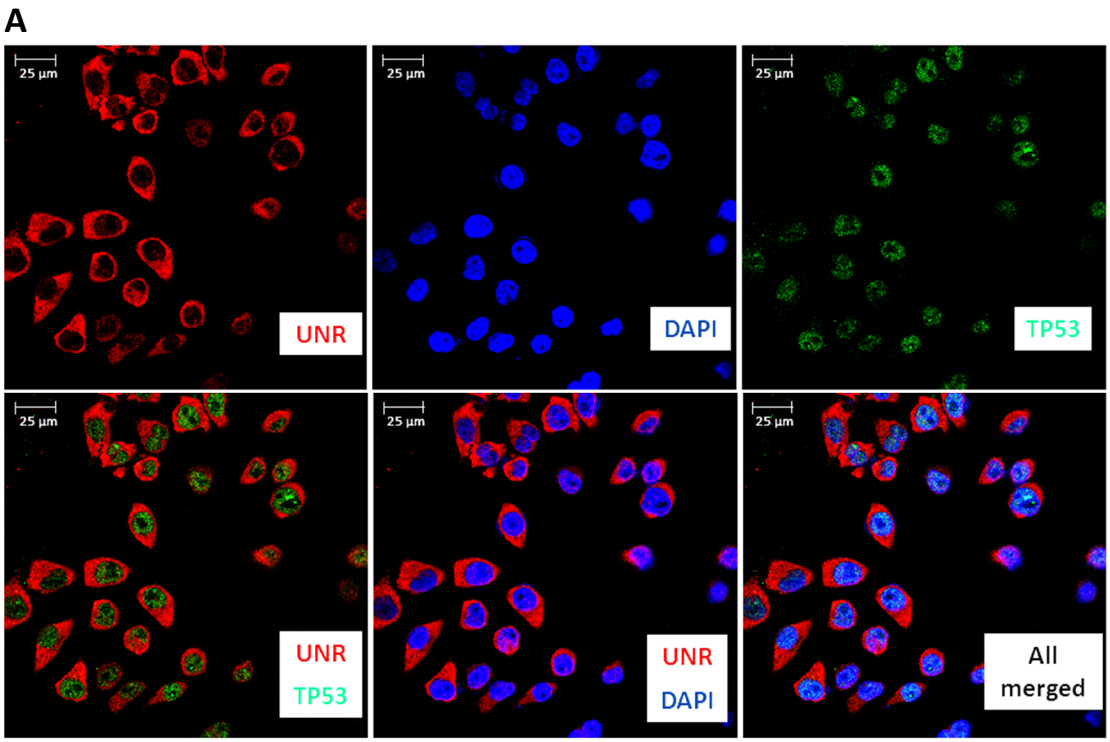


Figure 3.6: Confocal immunofluorescence microscopy images of non-arsenite treated HeLa cells (upper panel) or arsenite-stressed HeLa cells (lower panel) that were fixed and stained immediately after having been treated with 0.5 mM sodium arsenite in fresh DMEM (containing 10% FCS) or mock treated with sterile PBS in fresh DMEM (containing 10% FCS) for 30 minutes. The images show background staining from the fluorophore-conjugated secondary antibodies, as coloured, in the absence of primary antibodies. The scale bars are all 25 µm.



B

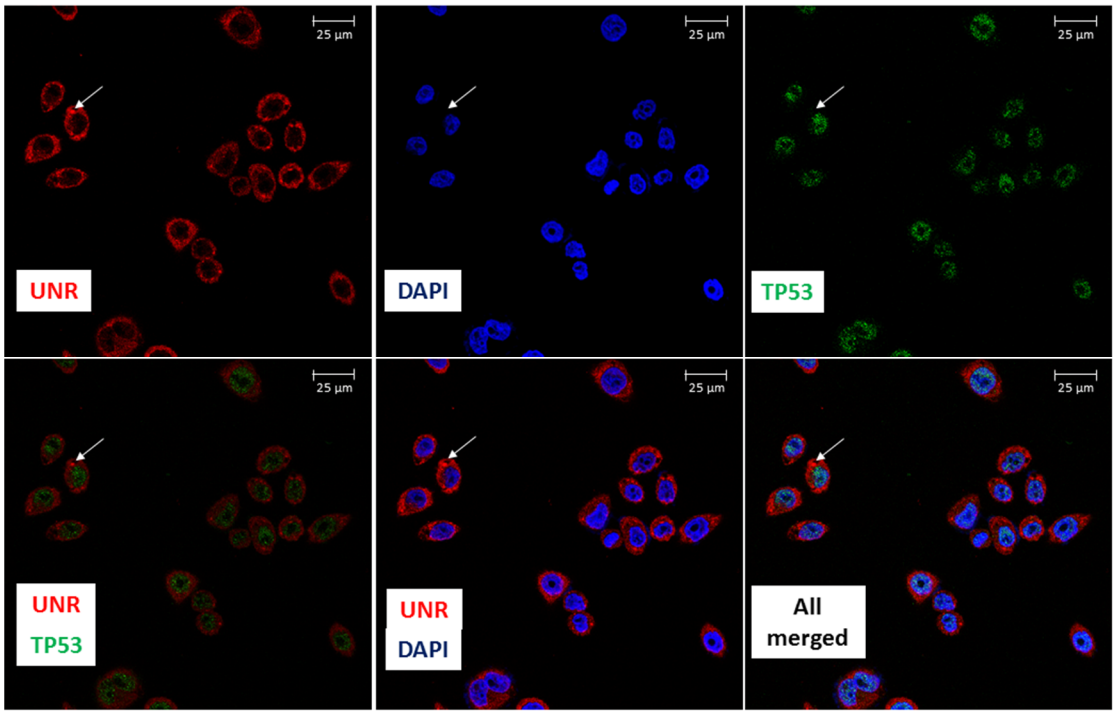


Figure 3.7: Confocal immunofluorescence microscopy images of non-arsenite treated HeLa cells (A) or arsenite-stressed HeLa cells (B) that were fixed and stained immediately after having been treated with 0.5 mM sodium arsenite in fresh DMEM (containing 10% FCS) or mock treated with sterile PBS in fresh DMEM (containing 10% FCS) for 30 minutes. The images show staining for UNR (red), TP53 (green), the DNA-binding dye DAPI (blue) or a combination of these (as stated). The white arrows in Figure 3.7B show UNR concentrated in a punctate structure. The scale bars are all 25 µm.

There was little observed background staining caused by the secondary antibodies alone (Figure 3.6).

It was observed that UNR had a predominantly cytoplasmic distribution in both unstressed (Figure 3.7A) and arsenite stressed (Figure 3.7B) HeLa cells. It was also noted that UNR was present in a number of punctate structures in arsenite stressed but not unstressed cells. An example of one of these structures is highlighted with arrows in Figure 3.7B. The structures could have been stress granules but it was not possible to state that with confidence. It was noted that TP53 did not appear to colocalize with UNR under either condition.

3.1.6 TP53 may colocalize with UNR in stress granules following arsenite stress for 1 or 2 hours

Seeing that UNR, but not TP53, migrated to stress granule-like structures following 30 minutes of arsenite stress, it was decided to widen the investigation to include the stress granule marker TIA1 to ascertain whether or not UNR and TIA1 colocalised within the same structures. It was also decided to look at longer time points (one and two hours) in order to see whether or not TP53 might colocalize with UNR over a longer time frame. HeLa cells were used as before and the time points are taken from the point of the administration of 1 mM sodium arsenite (zero minutes = arsenite was not added) (Figure 3.8). Previous results in the lab had suggested that the higher concentration of arsenite generated clearer images without having an additional detrimental effect on cell viability.

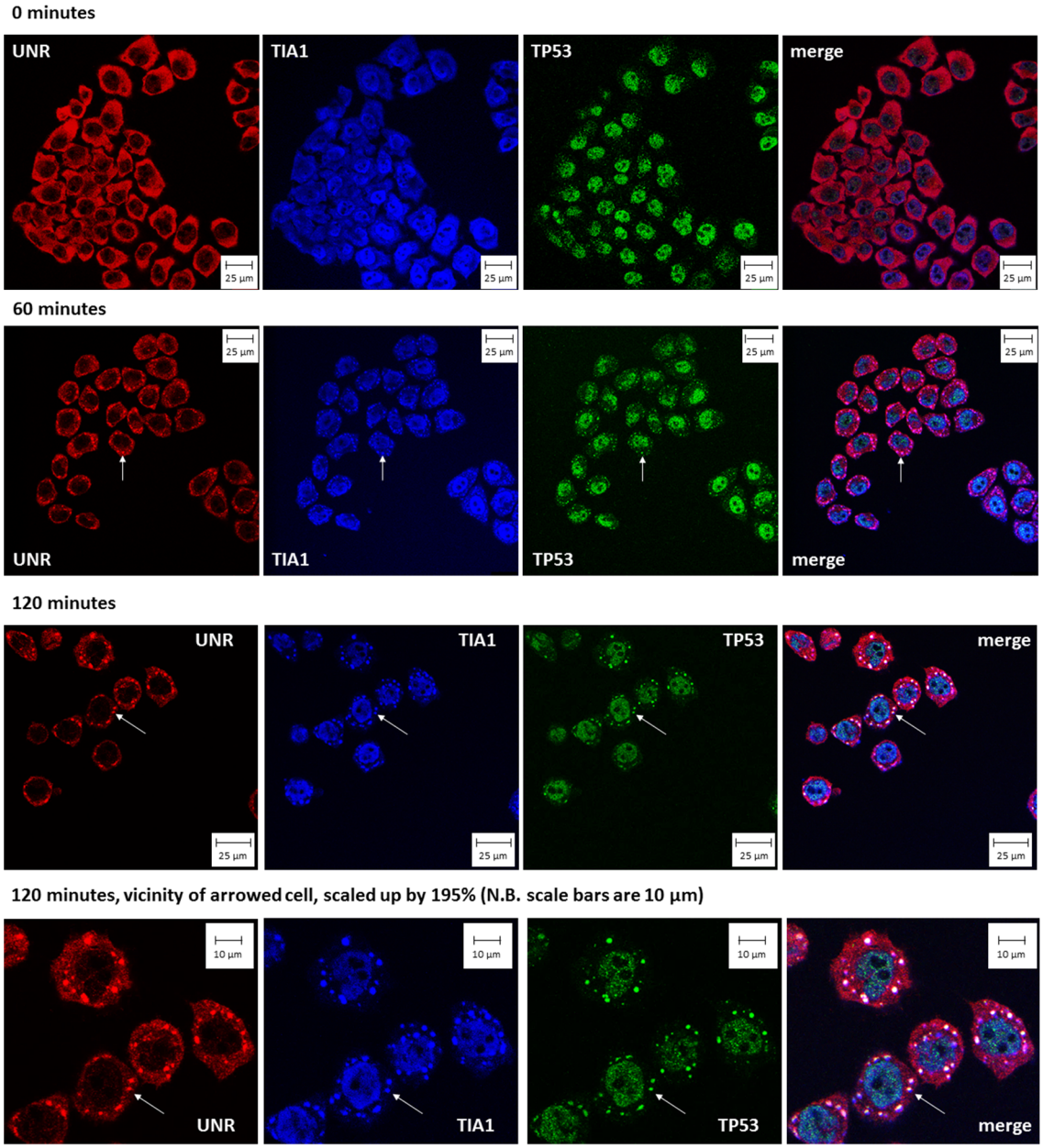


Figure 3.8: Confocal immunofluorescence microscopy images of non-arsenite treated HeLa cells (upper panel) or HeLa cells that were arsenite-stressed for 1 hour or 2 hours (as stated) and that were fixed and stained immediately after having been treated with 1 mM sodium arsenite in fresh DMEM (containing 10% FCS) or mock treated with sterile PBS in fresh DMEM (containing 10% FCS). The images show staining for UNR (red), TP53 (green), the stress granule marker TIA1 (blue) or a combination of all three (as stated). The white arrows at 1 and 2 hours show colocalisation of all three proteins into stress granules. The lower images are zoomed-in versions of a collection of cells from the 120 minute panels with 10 µm scale bars. All other scale bars are 25 µm.

UNR was shown to become locally concentrated in punctate structures by 60 minutes post-application of 1 mM sodium arsenite (Figure 3.8). The co-localisation of TIA1 and UNR in these punctate structures corroborates the hypothesis that the structures are stress granules (Figure 3.8). Furthermore, TP53 also appeared to co-localise with UNR and TIA1 in the structures after 60 minutes, with particularly clear co-localisation being apparent at 120 minutes (Figure 3.8). It was shown in the lab that UNR and TP53 can co-localise following stress with sodium arsenite in U2OS cells by IP-Western when pulling down with an antibody against TP53 (Ray, Ó Catnaigh and Anderson, manuscript in preparation). Confusingly, this interaction was never seen in the other direction (i.e. pulling down with an antibody against UNR and looking for TP53 by Western or in mass spectrometry data). These IF images corroborate the finding that UNR interacts with TP53 in stressed U2OS cells and this interaction will be the subject of ongoing investigation in the Anderson lab.

3.2 UNR can be immunoprecipitated and then detected by Western blotting

3.2.1 Immunoprecipitation of recombinant UNR

It was then decided to show that it was possible to immunoprecipitate UNR from IP buffer containing pure, though degraded, recombinant UNR protein.

To do this, 25 µl of Protein A-conjugated magnetic bead slurry was washed and placed in a blocking solution (5% BSA (w/v) in 400 µl of NT2) in a 1.5 ml tube. This was spun at 18rpm for one hour in a refrigerated room. Following this, 5 µg of an anti-UNR immunoglobulin raised in rabbit (see section 2.1.1), or a conspecific IgG was added and the solution was spun at 18rpm for four hours in a refrigerated room prior to washing. Finally, 10 ng of recombinant protein was dissolved in 100 µl of NT2 buffer. This was then added to a solution containing 857 µl of NT2, 40 µl of 0.5 M EDTA (at pH=8), 2 µl of VRC and 1 µl of 1 M DTT and the entire solution was added to the magnetic beads. The IP was then carried out overnight with spinning at 18rpm in a refrigerated room. The next day, the beads were washed and boiled in SDS loading buffer to dissociate the immunoprecipitated material. This was run on a 10% polyacrylamide gel prior to transferring the proteins to nitrocellulose membrane. The membrane was then blocked and probed with antibodies against UNR that were raised in goat. A HRP-conjugated anti-goat IgG secondary antibody was then applied and the membrane was ultimately developed using an enhanced chemiluminescence-based approach (Figure 3.9).

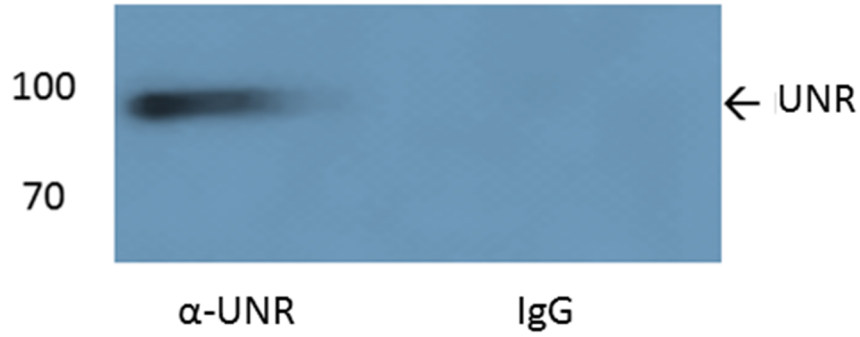


Figure 3.9: Result of an IP-Western probing for UNR in which an anti-UNR immunoglobulin raised in rabbit or a conspecific control IgG, as stated, was used to immunoprecipitate recombinant UNR protein from an IP buffer solution. The immunoprecipitated proteins were boiled in loading buffer and run on a 10% polyacrylamide gel prior to being transferred onto nitrocellulose membrane which was blocked, probed with an anti-UNR immunoglobulin raised in goat followed by a HRP-conjugated anti-goat IgG secondary antibody, and then developed using an enhanced chemiluminescence-based approach. The numbers on the left hand side of the image are approximate molecular masses (in kilodaltons) as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher).

Although the legacy antibody (see section 2.1.1) was able to detect multiple degradation products in the pure recombinant UNR Western (Figure 3.1), none were detected in the IP (Figure 3.9). This suggested that the immunoprecipitating antibody did not immunoprecipitate these or that the anti-UNR immunoglobulin raised in goat (see section 2.1.1) could not detect them by Western blot. Either suggestion would be surprising as the antibodies were both polyclonal. It had been decided not to run an input lane as the presence of UNR in the sample was known. In hindsight, it would have been interesting to prove that full length UNR was exclusively immunoprecipitated in the presence of multiple degradation products by including an input lane.

3.2.2 Immunoprecipitation of UNR from HeLa cell lysate

Carrying out an UNR IP-Western and probing for UNR with a different antibody showed that UNR could be pulled down from HeLa cell lysate (Figure 3.10).

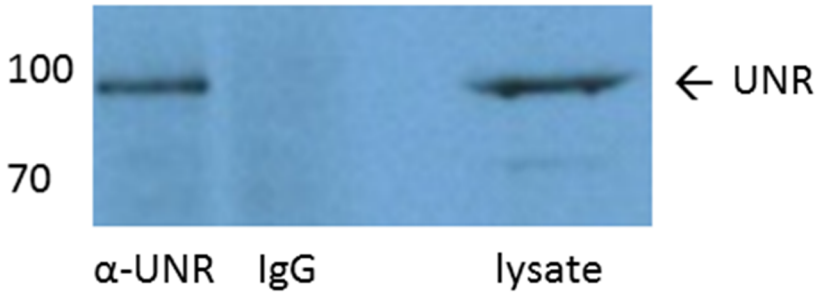


Figure 3.10: Result of an IP-Western probing for UNR in which an anti-UNR immunoglobulin raised in rabbit or a conspecific control IgG, as stated, was used to immunoprecipitate proteins from HeLa cell lysate. An input lane containing 5% of the total amount of lysate used in the experiment was loaded (labelled lysate). The immunoprecipitated proteins were boiled in loading buffer and run on a 10% polyacrylamide gel prior to being transferred onto nitrocellulose membrane which was blocked, probed with an anti-UNR immunoglobulin raised in goat followed by a HRP-conjugagted anti-goat IgG secondary antibody and then developed using an enhanced chemiluminescence-based approach. The numbers on the left hand side of the image are approximate molecular masses (in kilodaltons) as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher).

3.2.3 Co-immunoprecipitation of PABP with UNR

It has been recorded in the literature that UNR interacts with the PABP protein and additional interactions could be assumed from the pulldown of intact ribonucleoprotein complexes (RNPs). It was decided to show whether or not this protein could be detected by Western blot analysis from proteins immunoprecipitated with UNR (Figure 3.11).

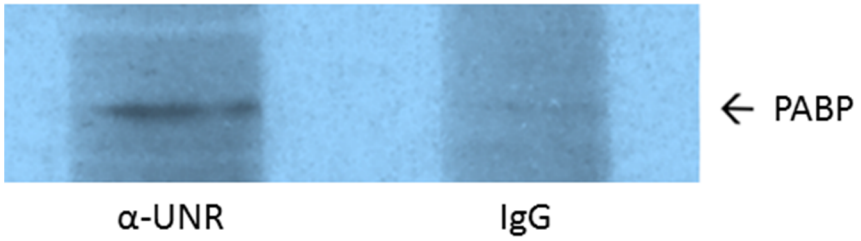


Figure 3.11: Result of an IP-Western probing for PABP in which an anti-UNR immunoglobulin raised in goat or a conspecific control IgG, as stated, was used to immunoprecipitate proteins from HeLa cell lysate. The immunoprecipitated proteins were boiled in loading buffer and run on a 10% polyacrylamide gel prior to being transferred onto nitrocellulose membrane which was blocked, probed with an anti-PABP immunoglobulin raised in rabbit followed by a HRP-conjugated anti-rabbit IgG secondary antibody and then developed using an enhanced chemiluminescence-based approach. This is the only IP recorded in this thesis that was carried out using the anti-UNR immunoglobulin raised in goat (Santa Cruz [sc-79292]; see section 2.1.1).

It was shown that UNR interacts with PABP (Figure 3.11). It was clear that the UNR pulldown sample included a protein at the size of PABP whereas the IgG pulldown had a much weaker band at that size.

3.2.4 Co-immunoprecipitation of the *PABP* transcript with UNR

It has also been stated in the literature that UNR interacts with the *PABP* transcript. It was also decided to show that this finding could be reproduced in the lab. This experiment was carried out with assistance from Dr Swagat Ray using cDNA he had generated previously from RNA he obtained from HeLa cell lysate and UNR/IgG IPs using the same lysate.

Briefly, 10pmol each of forward and reverse *PABP* primers were mixed with with nuclease-free water and Biomix Red (Bioline, to an end concentration of 1x). To 18 µl of this mixture was added 2 µl of test solution. The test solutions were cDNA from a UNR IP, cDNA from an IgG IP, a total cell cDNA positive control and a nuclease-free water negative control. The mixes were then subjected to amplification in a PCR machine with the following conditions:

1) Heat lid to 105°C
2) 30s at 94°C
3) 30s at 57°C
4) 30s at 72°C
5) Repeat steps 2-4 a further 29 times
6) Hold at 4°C until samples are removed

The cDNA samples subjected to electrophoresis on a 2% agarose gel containing ethidium bromide with a DNA ladder (the gel was run at 100 V / 60 mA for 90 minutes). The gel was then imaged in a UV box (Figure 3.12).

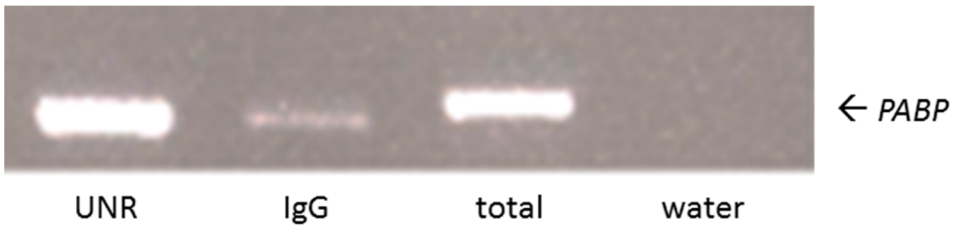


Figure 3.12: Polymerase chain reaction amplification of cDNA generated from the *PABP* transcript in HeLa cells. The lanes show cDNA amplified from *PABP* transcripts present in: complexes immunoprecipitated using an anti-UNR immunoglobulin (UNR), complexes immunoprecipitated using a control IgG (IgG), total cell RNA (total) and nuclease-free water (water). In each case, RNA was converted to cDNA prior to the PCR reaction being performed.

It seemed clear that UNR interacts with the *PABP* transcript as the band formed using the RNA from the UNR IP was more intense than that formed using the RNA from the IgG pulldown (Figure 3.12).

3.3 UNR distribution in the osteosarcoma cell lines U2OS and SaOS-2

3.3.1 UNR distribution in arsenite-stressed and unstressed U2OS and SaOS-2 cells by immunofluorescence microscopy

The confirmation in HeLa was considered strong enough evidence to proceed with initial IP-mass spectrometry experiments, initially with HeLa cells and subsequently with U2OS and SaOS-2 cells. As other cell types were to be used, it was decided to first check by immunofluorescence microscopy that the distribution of UNR was the same in U2OS (Figure 3.13) and SaOS-2 (Figure 3.14).

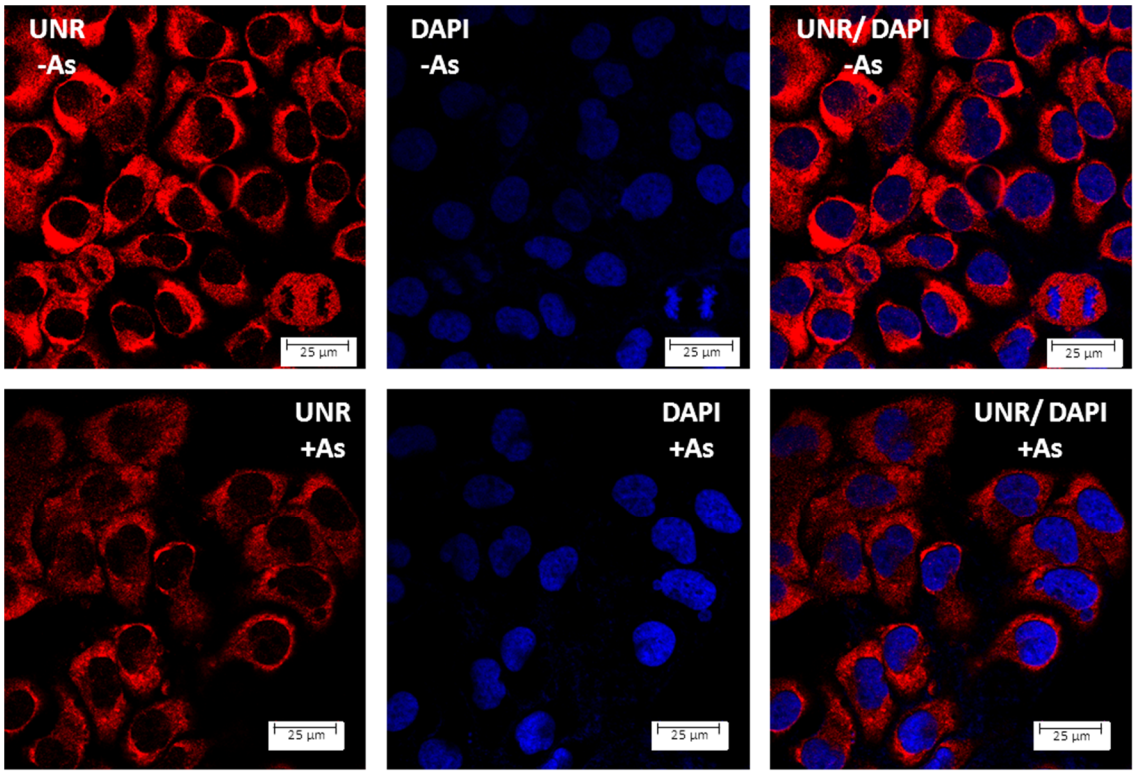


Figure 3.13: Confocal immunofluorescence microscopy images of non-arsenite treated U2OS cells (upper panel) or arsenite-stressed U2OS cells (lower panel) that were fixed and stained immediately after having been treated with 1 mM sodium arsenite in fresh DMEM (containing 10% [v/v] FCS) (lower panel) or mock treated with sterile PBS in fresh DMEM (containing 10% FCS [v/v]) (upper panel) for 1 hour. The images show staining for UNR (red), the DNA-binding dye DAPI (blue) or a combination of these (as stated). The scale bars are all 25 µm.

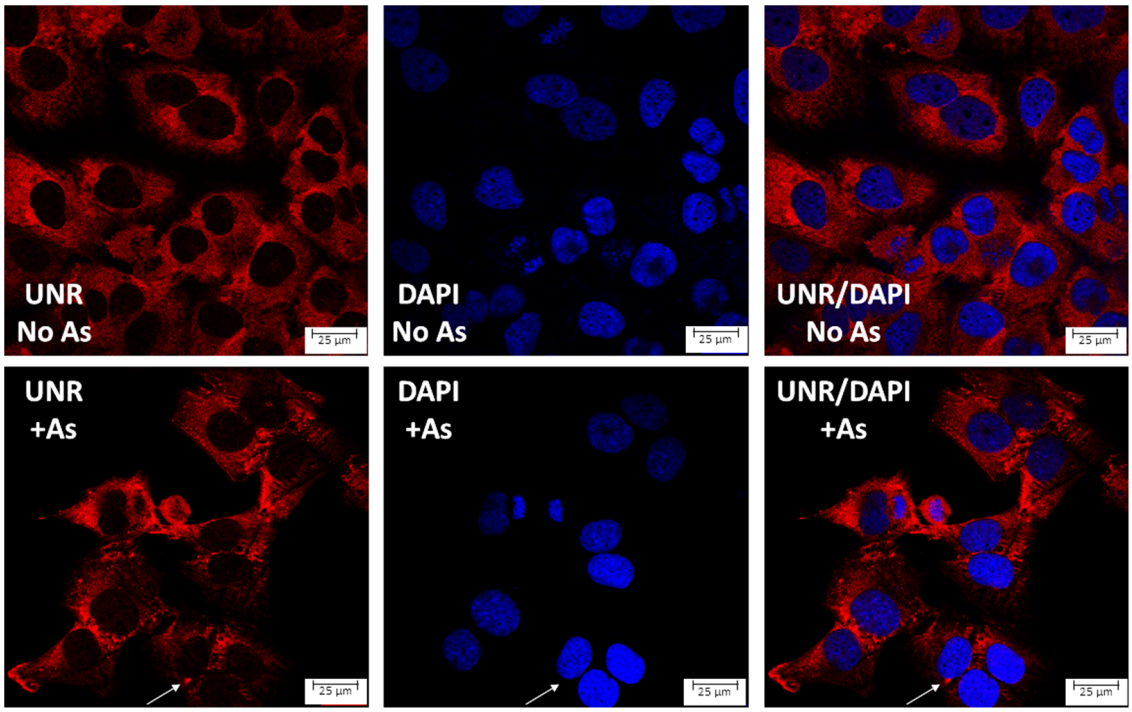


Figure 3.14: Confocal immunofluorescence microscopy images of non-arsenite treated SaOS-2 cells (upper panel) or arsenite-stressed SaOS-2 cells (lower panel) that were fixed and stained immediately after having been treated with 1 mM sodium arsenite in fresh DMEM (containing 10% [v/v] FCS) (lower panel) or mock treated with sterile PBS in fresh DMEM (containing 10% [v/v] FCS) (upper panel) for 1 hour. The images show staining for UNR (red), the DNA-binding dye DAPI (blue) or a combination of these (as stated). The white arrows in the lower panel show UNR concentrated in a punctate structure. The scale bars are all 25 µm.

This showed that UNR is predominantly cytoplasmic in U2OS and SaOS-2 cells. It also showed that the addition of sodium arsenite made both types of osteosarcoma cells to become less uniform (Figures 3.13, 3.14) in distribution and to become concentrated in areas of cell-cell contact, at least in SaOS-2 cells (Figure 3.14). This work was developed further in the lab and it was shown that arsenite stress causes UNR to localise to large stress granules in HeLa and SaOS-2 cells but that this was almost completely absent in U2OS cells. Smaller TIA1- but not UNR-containing structures were observed in arsenite-stressed U2OS cells. Interestingly, it was shown that UNR did become localised to large stress granules following arsenite stress in U2OS cells that had TP53 knocked down with siRNA (Ray, Ó Catnaigh and Anderson, manuscript in preparation). This led to the assumption that wild type TP53 is able to prevent the formation large UNR-containing stress granules.

3.3.2 UNR levels increase with increasing cell confluency in U2OS cells

Finally, it was decided to see if the cellular concentration of UNR exhibited the same inverse relationship with confluency that was observed in HeLa cells (Figure 3.4). It was decided to carry out a confluency course experiment using the wild type TP53-containing U2OS line. This allowed for both UNR and TP53 levels to be assessed to explore whether there could be a positive or negative correlation between levels of the two proteins. It was also decided to carry out the experiment using both unstressed and arsenite-stressed cells (Figure 3.15).

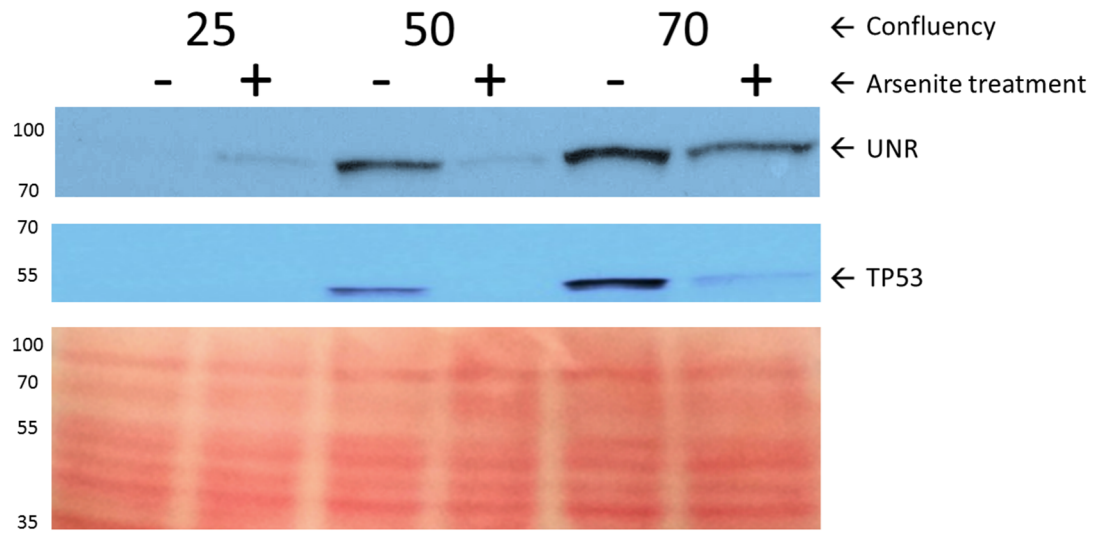


Figure 3.15: Western blot detection of UNR or TP53 (as stated) and Ponceau S staining of the pertinent regions of the membrane, by confluency at the point of harvesting U2OS cells, as stated. The cells were either treated with 1 mM of sodium arsenite ("+"), or a similar volume of sterile PBS ("-"), in fresh DMEM (containing 10% FCS) for 1 hour immediately prior to harvesting. 50 µg of total protein was loaded per lane. The numbers on the left hand side of the image are approximate molecular masses (in kilodaltons) as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher).

This showed the opposite relationship between UNR levels and cell confluency in unstressed U2OS to that observed in HeLa cells (Figures 3.4, 3.15). It also showed positive correlation between the levels of UNR and TP53 and that levels of both proteins were reduced at all confluencies following arsenite stress (with the exception of 25% confluency where little or no

UNR or TP53 was observed in either the plus or minus arsenite samples) (Figure 3.15). The reason for the difference in the confluency-UNR level relationships in HeLa and U2OS is currently unknown. The confluency-UNR level pattern in HeLa cells had been observed on multiple occasions (data not shown) and it was considered possible that the TP53 status of the cells could affect the relationship. It would be interesting to explore this possibility further. Likewise, it would be interesting to explore how the expression of wild type TP53 in U2OS cells is able to prevent the formation of large stress granules following arsenite stress. It was confusing that knocking down TP53 allowed the formation of large stress granules following arsenite stress in U2OS cells (Ray, Ó Catnaigh and Anderson, manuscript in preparation). This was because arsenite stress itself was shown to reduce the cellular level of TP53 (Figure 3.15) without the formation of large stress granules (Ray, Ó Catnaigh and Anderson, manuscript in preparation, Figure 3.13).

Having seen that UNR is expressed in all three cell types to be used and demonstrated that it is possible to pull it down with antibodies available to the group, it was decided to proceed to carry out large-scale experiments to discover novel UNR-interacting proteins and RNAs.

4 Identification of UNR-interacting proteins

As explored in Chapter 1, much remains to be elucidated with respect to the function of UNR in mammalian cells. In order to build upon the current knowledge of UNR function, it was decided to search for novel UNR-interacting proteins by carrying out ribonucleoprotein immunoprecipitations with an antibody against UNR, or a conspecific control IgG, followed by subjecting the immunoprecipitated samples to analysis by mass spectrometry.

Dr Swagat Ray kindly assisted with the ribonucleoprotein immunoprecipitations using the HeLa (see section 4.3) and U2OS (see section 4.5) lysates. In particular, he carried out the post-IP Dynabead washes for these cell types prior to the samples being split for mass spectrometry or RIP-Seq (see section 2.2.9).

4.1 RIP-mass spec

4.1.1 Introduction to RIP

In order to explore the cellular functions of UNR, it was first necessary to build upon the known repertoire of UNR-interacting proteins and RNAs that are recorded in the literature. Given that it had previously been shown that the presence or absence of UNR appeared to lead to differences in the transcriptome (Dormoy-Raclet, 2005; Elatmani *et al.*, 2011) and that UNR has been recorded in the literature as binding to a variety of proteins and transcripts (reviewed in Ray et al. 2015), it was decided to take a methodical look at the proteins and RNA species to which it can bind. In order to obtain UNR in natural ribonucleoprotein (RNP) complexes, as well as UNR in specific complexes with proteins or RNAs alone, a series of ribonucleoprotein immunoprecipitations (RIP) were undertaken (Keene *et al.*, 2006). RIPs have an advantage over standard immunoprecipitations in that they can allow for the same samples to be used for both protein and RNA analysis.

Briefly, cells of the given cell type were grown in multiple 15 cm plates (i.e. circular plates with a diameter of 15 cm and a surface area of 56.25π cm^2). One hour prior to harvesting, half the plates were treated with sodium arsenite at an end concentration of 1 mM in fresh DMEM containing 10% FBS. The other half were fed with fresh DMEM containing 10% FBS with the same volume of sterile PBS as sodium arsenite as a mock treatment. This was repeated on two further

occasions for each cell type, so as to generate three biological repeats. After harvesting, each lysate was assayed for protein concentration, separated into two and RIPs were carried out using an antibody against UNR or a nonspecific control IgG raised in the same animal (Figure 4.1). The samples were then separated into two, with 80% being processed to extract RNA for RNA sequencing. The remaining 20% was processed to prepare proteins for mass spectrometry. Initial trial runs were carried out using unstressed HeLa cells. In these cases, less material was used and none was set aside for RNA purification (see below).

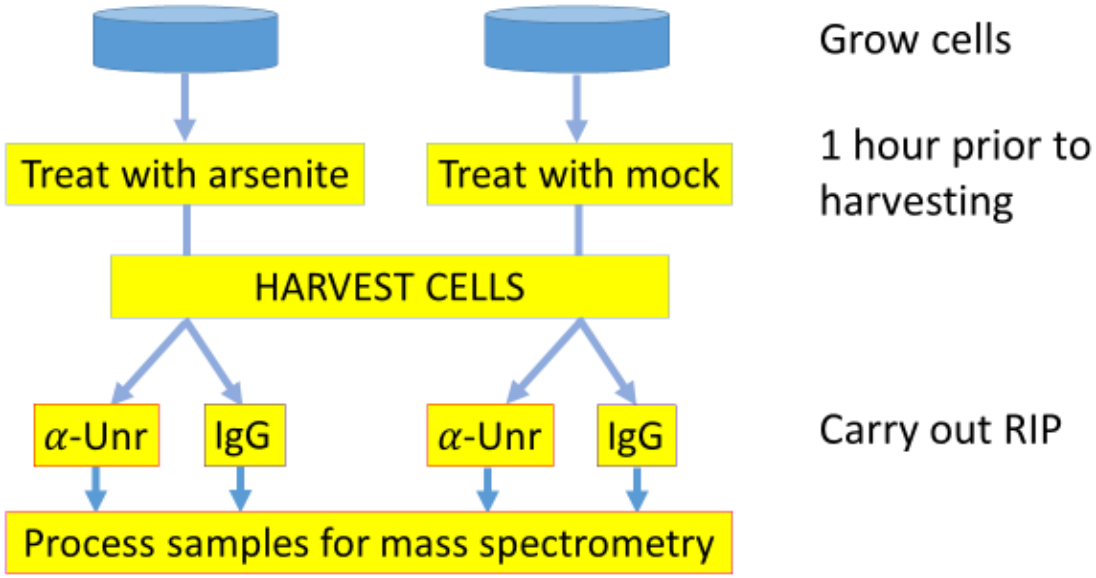


Figure 4.1: Schematic flow chart of main steps involved in the preparation of samples for mass spectrometry. α-Unr refers to using an antibody against UNR for the RIP whereas IgG refers to the use of a conspecific nonspecific control antibody. Further details on the individual steps can be found in the text.

4.1.2 Introduction to RIP-mass spectrometry

In order to discover a selection of direct and indirect UNR-interacting proteins, it was decided to subject the proteinaceous portion of the pulldown sample to analysis by mass spectrometry (Aebersold & Mann, 2003).

Briefly, Protein A-bound magnetic beads were used to bind to the immunoprecipitating immunoglobulins that were, in turn, used either to pull down the protein of interest or act as a

negative control. The samples were then washed and processed, prior to being treated with modified pig trypsin that cleaves proteins at predictable sites. This generated mixes of peptides that were separated by nano-liquid chromatography before being ionised (electrospray ionisation) over time and analysed by an Ultimate 3000/Orbitrap Fusion mass spectrometer (Thermo Scientific). This takes a series of ionised peptides over a given time interval and generates an MS1 spectrum, internal software then selects some individual peptides for fragmentation and further analysis, yielding MS/MS spectra (Aebersold & Mann, 2003). By the end of the run, the amassed data can be used to reconstruct the likely components of the original mixture based on algorithms that piece together the predicted peptides into the most likely combination of proteins that could have generated the mixture.

4.2 Choice of mass spectrometry sample preparation method

4.2.1 IP-mass spec basics

Immunoprecipitations, be they RIPs or standard IPs, are used to pull down a bait protein bound to a mixture of other proteins (Figure 4.2).

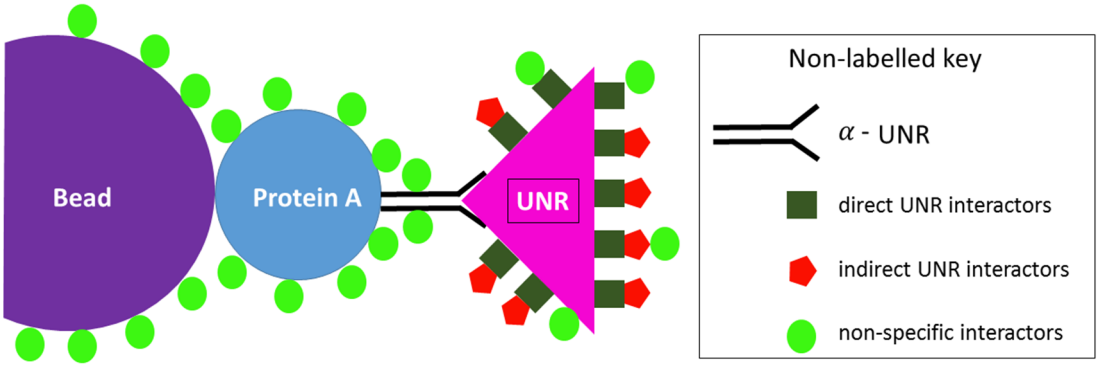


Figure 4.2: Schematic diagram of a UNR immunoprecipitation. 'Protein A' is bound to suitable beads and then used to bind to an anti-UNR IgG. The antibody then pulls down its bait protein and an array of associated molecules. Non-specific interactors will also be pulled down and some will remain after washing. N.B. the term 'non-specific interactor' is used to refer to molecules not binding specifically to the bait protein or other molecules in complex with it; these interactions may actually be quite strong.

4.2.2 On-bead sample preparation compared with SDS-PAGE gel slice method

Initial experiments were carried out to compare two different methods of sample preparation for subsequent mass spectrometry experiments. The two methods considered were termed the 'gel slice' method and the 'on-bead' method, with the latter also being termed 'in solution'. Unstressed HeLa cells were used for the trial experiments.

4.2.3 Additional steps are required for the gel slice method over the on-bead method

Whereas the IP samples are processed immediately for the on-bead tryptic digest method, for the gel slice method the samples must first be stripped from the beads, separated by mass in an SDS-PAGE gel using electrophoresis and the gel must then be cut before the individual gel pieces can be prepared for mass spectrometry (Figure 4.3).

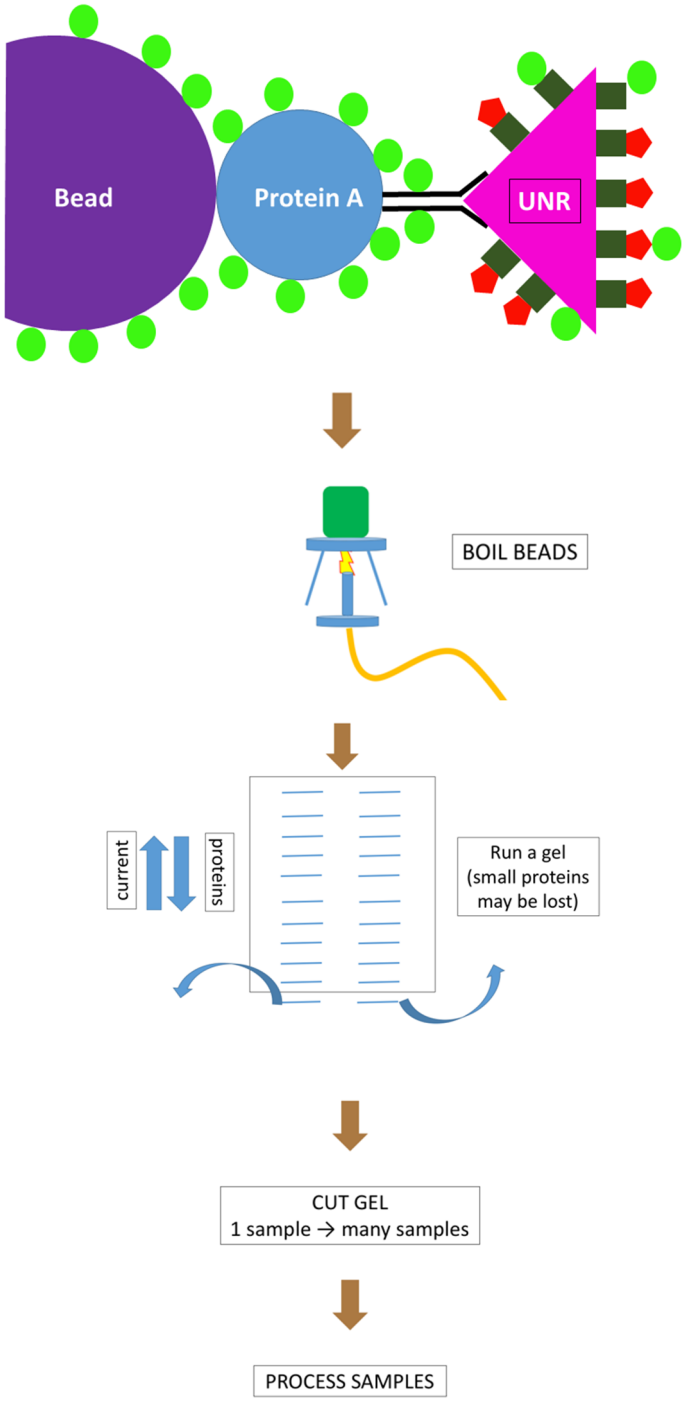


Figure 4.3: Extra processing steps are required for the gel slice method over the on-bead method. This diagram is purely schematic and neither to scale nor representative of the actual laboratory equipment used to carry out each step. Also see key from Figure 4.2.

4.2.4 Advantages and disadvantages of the on-bead and gel slice methods

Of the two methods, the on-bead method had advantages in terms of being more quantitative, requiring less processing time and being less technically challenging. All proteins can be submitted without losses due to small proteins running off gels. Another reason for the gel slice method being less quantitative is due to almost unavoidable small differences in where gels are cut between samples. This can alter the distribution of proteins between slices which, in turn, can alter their probability of being detected as their peptides will be competing with peptides from a different set of proteins.

On the positive side for the gel slice method, having a rough idea of the where a gel has been cut can offer an additional layer of confidence that a hit is genuine by cross referencing its predicted size to the gel slices in which it was detected (Figure 4.4).

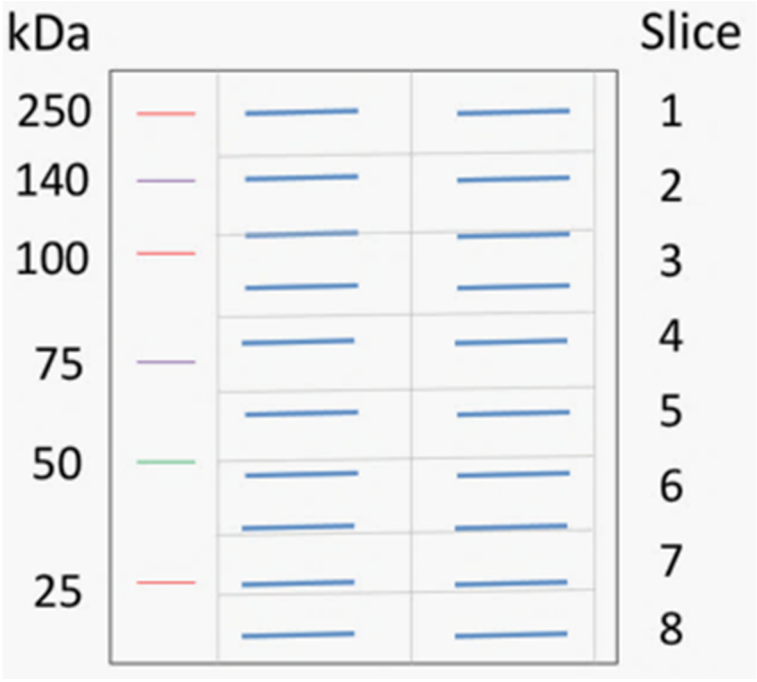


Figure 4.4: Schematic diagram of a gel that has been cut relative to hypothetical coloured marker bands. The faint grey lines refer to the cut points of the gel, such that the slices numbered to the right of the figure are generated. The blue lines signify a selection of hypothetical proteins separated by size by passing an electrical current through the gel.

Using the example above, should a putative protein be 50 kDa in mass and it was found to have no peptides in slices 1-4 but lots in slices 5 and 6 and fewer in 7 and 8, the data would be in keeping with a hypothesis that the gel was sliced through the protein and that some degradation products were found further down the gel. If a protein that was predicted to be 175 kDa were to have the same distribution, that may imply that the hit was a false positive. Such information would not be available should the on-bead method be used.

4.2.5 Results for trial run using the on-bead method

For the on-bead method, HeLa samples were prepared as stated previously. An initial injection of 2 µl per sample was run for 30 minutes, followed by a run of 5 µl for three hours. A final run of 20 µl for three hours was then carried out and the data were merged in Scaffold after Mascot searches were carried out, as per [supplementary data]. The RIPs were carried out using proteins that were estimated to fall within the suggested range of 2-5 mg total protein per IP (Keene *et al.*, 2006). As the samples were ultimately resuspended in 50 µl, this implied that 2 µl corresponded to 40-100 µg of starting material, 5 µl corresponded to 100-250 µg of starting material and 20 µl corresponded to 400-1000 µg of starting material.

Within Scaffold, the data from the three runs were merged for each of the two conditions (IgG and UNR) and protein quantification was set to total spectra. Fisher's Exact test was used protein-wise and the level of significance was set to p=0.05, modified for multiple testing using the Hochberg-Benjamini correction. The correction lowered the cut off level of significance to p=0.01623.

With the minimum number of peptides required to identify a protein set at two, the peptide threshold set at 95% and the protein threshold at 99%, 228 proteins were detected. In terms of presence/absence, 34 proteins were detected in the IgG samples only, 64 in the UNR samples only and 130 proteins in both samples. Using Scaffold's inbuilt Quantitative Profile function, 41 proteins were detected as significantly higher in the IgG samples, 32 proteins as higher in the UNR samples and 155 proteins as not significantly higher in either sample (Figure 4.5). It was noted that a proportion of the proteins identified were either keratins or non-human.

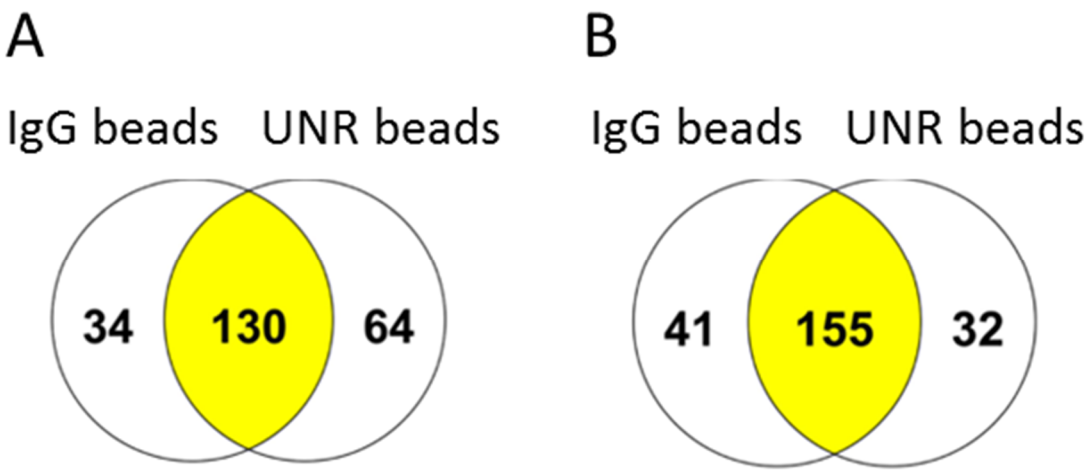


Figure 4.5: Venn diagrams summarising the results of the on-bead HeLa sample preparation test run. The numbers depict the number of proteins detected by Scaffold under the given thresholds, either in presence/absence (A) or by quantitative profile (B). The yellow shading depicts the number of proteins present in both samples (A) or not significantly over represented in either sample (B). Significance in this respect is assessed on whether or not the Fisher's exact test p-value for the protein is less than the adjusted p-value stated in the text ($p<0.01623$). The unshaded portions contain the number of proteins only present in the stated condition (A), or significantly overrepresented in the stated condition (B). The threshold for peptide identity was 95%, the threshold for protein identity was 99% with a minimum of two different peptides.

The volcano plot function in Scaffold graphically presents the log2 fold change between two conditions on the x-axis against the negative log10 of the p-value of a statistical test, in this case the Fisher exact test (Figure 4.6).

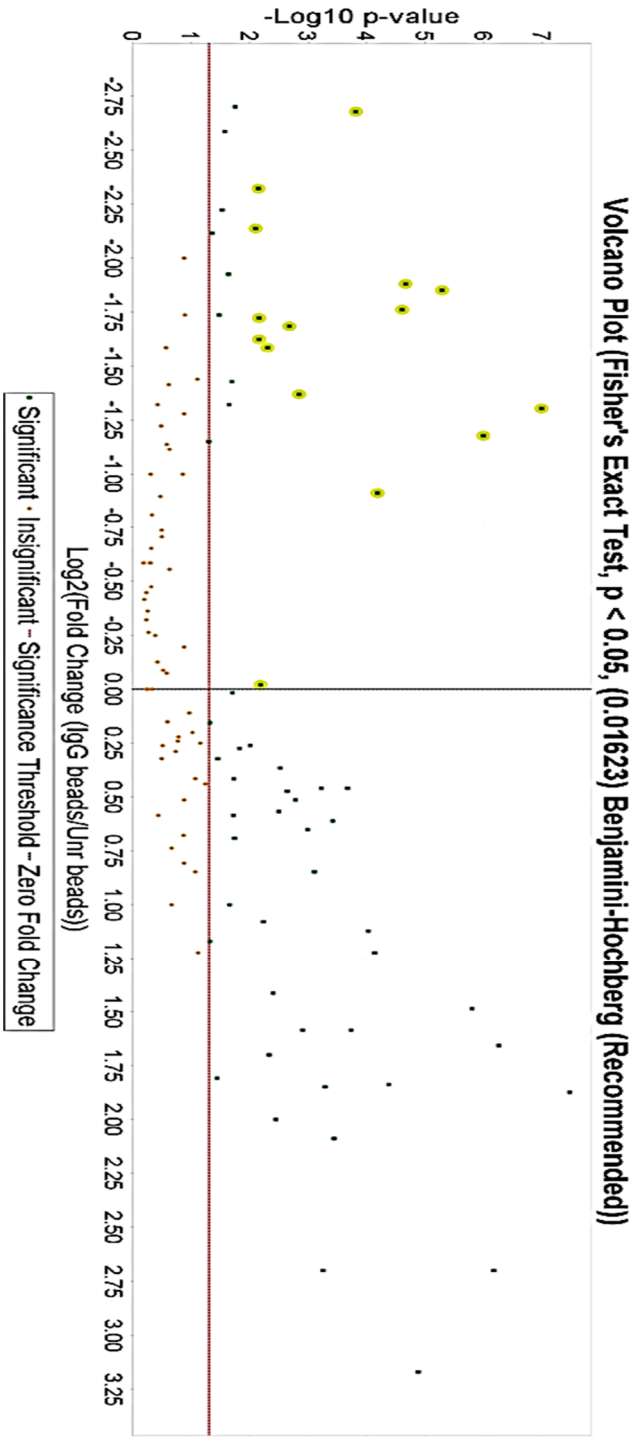


Figure 4.6: Scaffold volcano plot of the on-bead HeLa sample preparation test run. The log2 of the fold change between IgG and UNR is shown on the x-axis with a dark line at zero fold change. The y-axis plots the negative log10 of the Fisher exact test p-value for a given protein. The red line signifies the selected, though unadjusted, cut-off point for statistical significance (p=0.05). Orange dots beneath the red line have unadjusted p-values over 0.05. The green dots above the line have p-values at or below 0.05. The yellow circles highlighting certain green dots indicate proteins that pass the multiple testing correction for being over-represented in the UNR pulldown samples. Figure reproduced from Scaffold.

Several of the top hits were keratins and there were more hits for IgG than UNR. Positively, UNR and UNRIP were both present and considered significant (Figures 4.7A and 4.7B). There were also several proteins associated with RNA metabolism as well as a number of unexpected hits, including Lim domain binding protein 1 (LDB1) and single stranded DNA binding protein 3 (SSBP3) (Figures 4.7C and 4.7D).

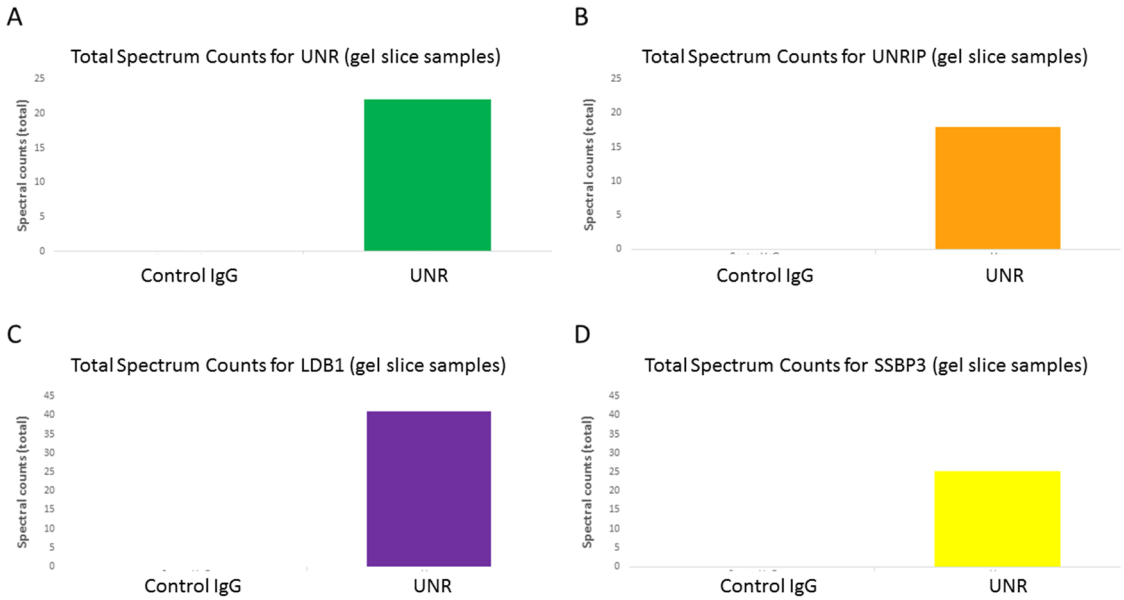


Figure 4.7: Number of spectra for specified proteins from normalised total spectra for IgG and UNR using the on-bead trial run data, as stated: UNR (A), UNRIP (B), LCB1 (C), SSBP3 (D). The scale for A and B is 0 – 25 spectral counts and the scale for C and D is 0 – 45 spectral counts. The counts were: UNR=22, UNRIP=18, LDB1=41, SSBP3=25.

4.2.6 Results for trial run using the gel slice method

For the gel slice method, HeLa samples were prepared as stated previously. RIP samples were boiled in loading buffer and separated by gel electrophoresis. The UNR and IgG sample lanes were then cut into 7 pieces arranged by size (first slice containing the largest proteins) and processed for further mass spectrometric analysis. 20 µl per sample was then run for 60 minutes. The same amount of total protein was used for the RIPs as for the on-bead method (2-5 mg per RIP). The samples were boiled in approximately 30 µl of loading buffer. Therefore, excluding sample concentration, due to buffer evaporation, approximately 1.3-3.3 mg of starting material were used to generate the samples loaded on each lane on the gel.

The data from the seven slices was then recombined into a single file and analysed by Fisher's exact test as above. The Hochberg-Benjamini correction lowered the cut off level of significance to p=0.01353.

With the minimum number of peptides required to identify a protein set at two, the peptide threshold set at 95% and the protein threshold at 99%, 510 proteins were detected. In terms of presence/absence, 292 proteins were detected in the IgG samples only, 67 in the UNR samples only and 151 proteins in both samples. Using Scaffold's inbuilt Quantitative Profile function, 87 proteins were detected as significantly higher in the IgG samples, 50 proteins were significantly higher in the UNR samples and 373 proteins were not significantly higher in either sample (Figure 4.8). It was noted that a proportion of the proteins identified were either keratins or non-human.

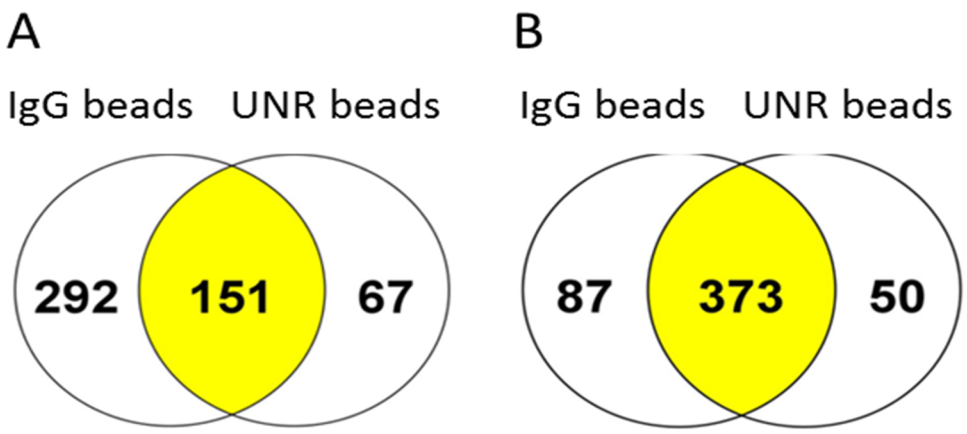


Figure 4.8: Venn diagrams summarising the results of the gel slice HeLa sample preparation test run. The numbers depict the number of proteins detected by Scaffold under the given thresholds, either in presence/absence (A) or by quantitative profile (B). The yellow shading depicts the number of proteins present in both samples (A) or not significantly over represented in either sample (B). Significance in this respect is assessed on whether or not the Fisher's exact test p-value for the protein is less than the adjusted p-value stated in the text ($p<0.01353$). The unshaded portions contain the number of proteins only present in the stated condition (A), or significantly overrepresented in the stated condition (B). The threshold for peptide identity was 95%, the threshold for protein identity was 99% with a minimum of two different peptides. Figure reproduced from Scaffold.

The volcano plot for the gel slice method, highlighting the proteins that passed the multiple testing correction, is shown below (Figure 4.9).

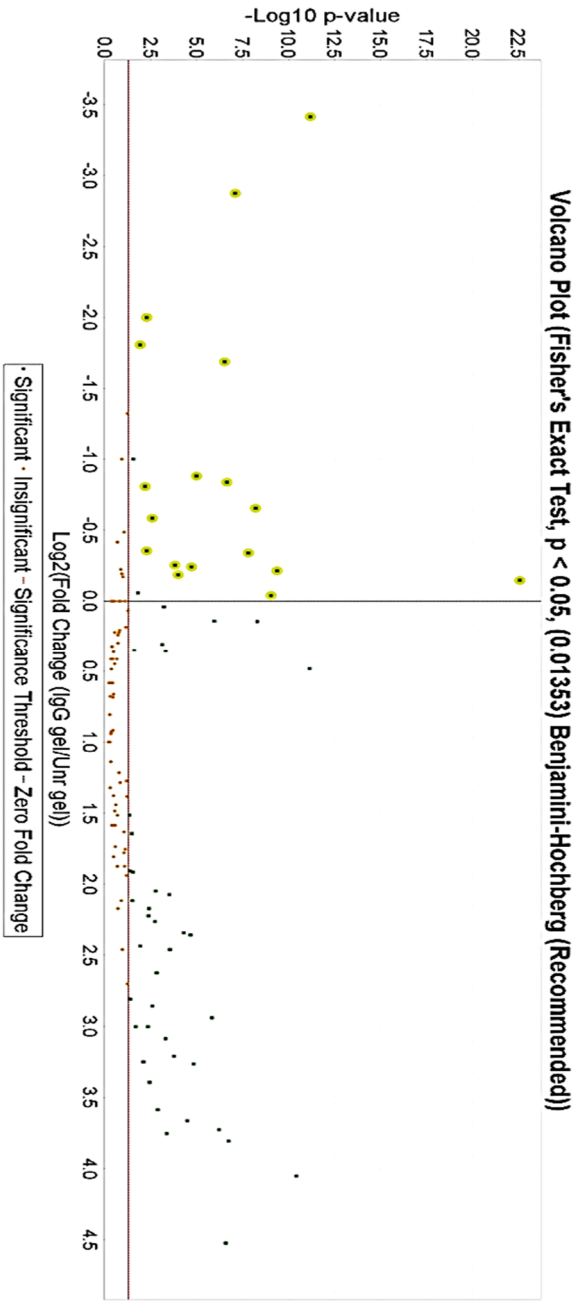


Figure 4.9: Scaffold volcano plot of the gel slice HeLa sample preparation test run. The log2 of the fold change between IgG and UNR is shown on the x-axis with a dark line at zero fold change. The y-axis plots the negative log10 of the Fisher exact test p-value for a given protein. The red line signifies the selected, though unadjusted, cut-off point for statistical significance (p=0.05). Orange dots beneath the red line have unadjusted p-values over 0.05. The green dots above the line have p-values at or below 0.05. The yellow circles highlighting certain green dots indicate proteins that pass the multiple testing correction for being over-represented in the UNR pulldown samples. Figure reproduced from Scaffold.

As with the on-bead method, several hits were either keratin or non-human. It was noted that there was some cross-over between the two sets of hits. For example, the selection of top hits shown in Figure 4.7 were also top hits by the gel slice method (Figure 4.10).

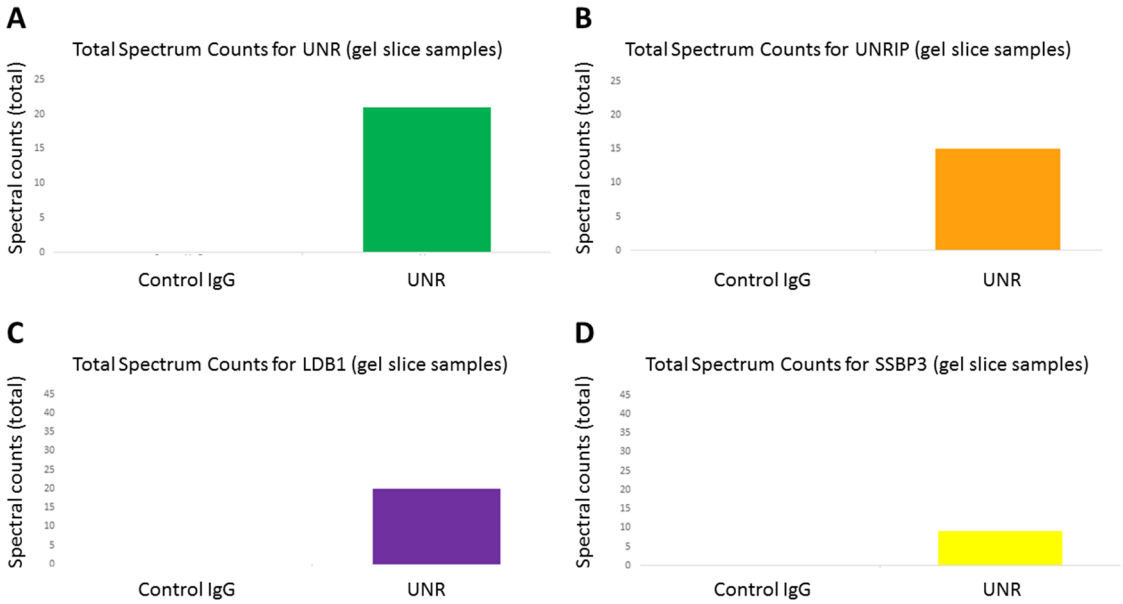


Figure 4.10: Number of spectra for specified proteins from normalised total spectra for IgG and UNR using the gel slice trial run data, as stated: UNR (A), UNRIP (B), LCB1 (C), SSBP3 (D). The scale for A and B is 0 – 25 spectral counts and the scale for C and D is 0 – 45 spectral counts. The counts were: UNR=21, UNRIP=15, LDB1=20, SSBP3=9. The scales used here are the same as those used in Figure 4.7.

4.2.7 Comparison of the results from the trial runs

The UNR-enriched quantitative hits flagged up by Scaffold for the different methods (using the parameters in the text) were considered, together with lists of proteins present in the UNR data but not in the IgG data for the two methods. The data for the two methods were combined in the following ways:

1) On-bead quantitative hits with gel slice quantitative hits
2) On-bead presence/absence data with gel slice presence/absence data
3) On-bead quantitative hits with on-bead presence/absence data
4) Gel slice quantitative hits with gel slice presence/absence data

In each case, duplicated Accession numbers were highlighted. In the case of 3 and 4, one copy of the duplicated values was removed and then the remaining proteins in each case were combined in a separate file (5). Duplicates within file 5 were then highlighted. The duplicated Accession numbers from (1), (2) and (5) were then tabulated and any keratins and non-human proteins were removed. Finally, only one copy of any duplicated values was retained (Table 4.1).

Table 4.1: Proteins statistically over-represented in UNR data over IgG data and/or only present in UNR data by both methods. Accession numbers referring to keratins and non-human proteins were removed, as were extra copies of duplicates

Accession	Protein Name
Q86U70	LIM domain-binding protein 1 (GN=LDB1)
O96028	Histone-lysine N-methyltransferase NSD2 (GN=WHSC1)
G5E9Q2	Cold shock domain containing E1, RNA-binding, isoform CRA_d (GN=CSDE1)
Q9BWW4	Single-stranded DNA-binding protein 3 (GN=SSBP3)
Q9UHX1	Poly(U)-binding-splicing factor PUF60 (GN=PUF60)
Q13435	Splicing factor 3B subunit 2 (GN=SF3B2)
A8K963	cDNA FLJ77516, highly similar to Homo sapiens LIM domain only 4 (LMO4), mRNA
Q5W009	RNA binding motif protein 17, isoform CRA_a (GN=RBM17)
O75376	Nuclear receptor corepressor 1 (GN=NCOR1)
P0DI83	Ras-related protein RAB34, isoform NARR (GN=RAB34)
B7Z475	cDNA FLJ55712, highly similar to F-box-like/WD repeat protein TBL1XR1
B4E2Z3	cDNA FLJ54090, highly similar to 4F2 cell-surface antigen heavy chain
E7EMC7	Sequestosome-1 (GN=SQSTM1)
F5GWP8	Junction plakoglobin (GN=JUP)
A8K9P0	cDNA FLJ78413, highly similar to Homo sapiens albumin, mRNA
P49756	RNA-binding protein 25 (GN=RBM25)
B3KWX7	cDNA FLJ44170 fis, clone THYMU2035319, highly similar to RNA-binding region-containing protein 2

It was considered reassuring that the bait protein was considered over-represented in both cases. This gave confidence that the RIPs and mass spec both worked. Both methods also allowed the mass spectrometry machine to detect a selection of peptides from other proteins in such quantities that Scaffold was able to flag up a significant over-representation of them in the UNR samples. The presence of keratins in the data was unwelcome but to be expected. Likewise, non-human proteins such as bovine serum albumin were used in the RIPs and their presence was also not unexpected. It was noted that a strange Accession number was assigned

to UNR (G5E9Q2) and UNRIP had different Accession numbers assigned to it between the two methods (Q9Y3F4 and B4DNJ6), explaining why it was excluded from Table 4.1. It should be noted that a Uniprot search for B4DNJ6 redirects to Q9Y3F4. Likewise, G5E9Q2 redirects to O75534. It was considered likely that these anomalies were down to the use of old Mascot searches. Unfortunately, as these were carried out by mass spectrometry technicians no longer employed by the University of Warwick, it is not possible to ascertain retrospectively if there were any non-documented differences in the methodology used by different proteomics technicians.

It was noted that more overall hits were detected for the gel slice method over the on-bead method. For this reason, two interesting proteins flagged up by the gel slice method were searched for within the on-bead data.

Due to the possible link between UNR and cancer, multiple myeloma tumor-associated protein 2 (MMTAG2) was considered (Figure 4.11). This protein was flagged as both significantly over-represented in the UNR data and also only present in the UNR data for the gel slice method. It was not significant for the on-bead method although it was detected.

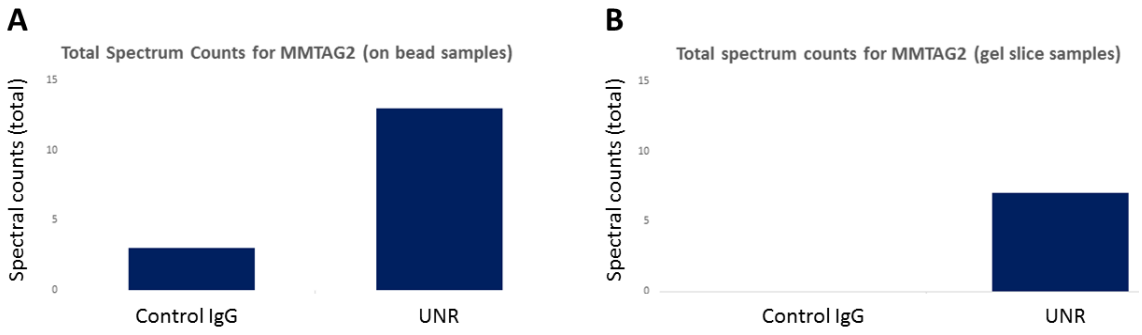


Figure 4.11: Number of MMTAG2 spectra for IgG and UNR by the on-bead method (A) and the gel slice method (B). The scales for A and B are both 0 – 15 spectral counts. The counts were - on-bead: IgG=3, UNR=13; gel slice: IgG=0, UNR=7.

Whilst MMTAG2 was not present in the IgG data for the gel slice method, the number of total spectra for it was almost twice as high in the on-bead method (13 versus 7). The detection of 3 MMTAG2 spectra in the on-bead IgG data changed the Fisher's exact test p-value for MMTAG2 such that it no longer passed the multiple testing correction.

This draws attention to the problem of looking at single sets of data without repeats. Biologically, pulling down 3 lots of a protein with IgG to 13 lots with an antibody against a protein of interest implies that the protein interacts in some way with the protein of interest. On the other hand, there were 361 spectra with IgG for keratin, type II cytoskeletal 1 (KRT1) in the gel slice data and 399 spectra for UNR and those data yielded a Fisher's exact test p-value of 2.6×10^{-23} even though it was very close to the zero fold change line on the volcano plot (Figure 4.9). An alternative would have been to use fold changes but that method also has shortcomings with regards to small absolute values (e.g. the change between 1 and 2 is more likely to be down to chance than a change between 100 and 200, even though the ratios would be the same). It also yields regular infinite values that need to be checked individually to see if they were of the order of 1 versus 0 or 100 versus 0 as the former is more likely to be a false positive whereas the latter is likely to be a true positive. As the purpose of this section was merely to compare two methods rather than making any biological conclusions, it was decided to choose the option that avoided infinities.

It was then decided to investigate 78 kDa glucose-regulated protein (HSPA5) because of recent literature linking UNR to diabetes, implying that UNR transcripts are downregulated more than 6 fold in patients with excessively high glucose levels over the same patients following treatment

that returned their blood glucose levels to lower and, in most cases, normal levels (Figure 4.12) (Xavier *et al.*, 2014). This protein was flagged as significant by the gel slice method but not by the on-bead method, although it was detected in both.

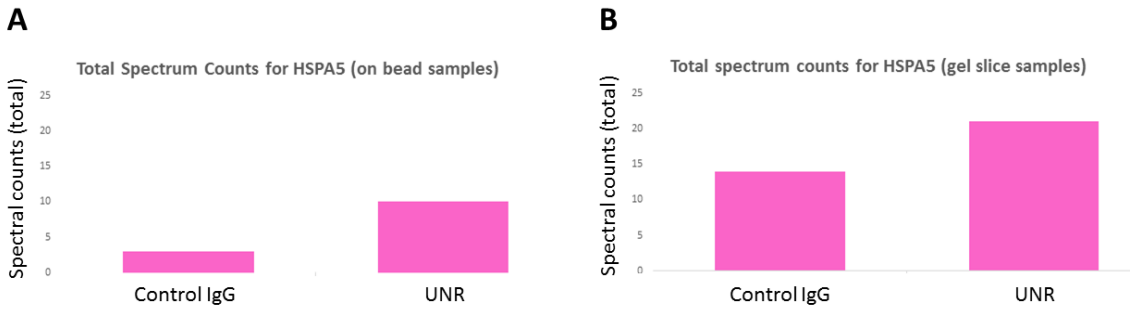


Figure 4.12: Number of HSPA5 spectra for IgG and UNR by the on-bead method (A) and the gel slice method (B). The scales for A and B are both 0 – 25 spectral counts. The counts were - on-bead: IgG=3, UNR=10; gel slice: IgG=14, UNR=21.

This case shows a different problem to that of detecting MMTAG2 as a hit. Akin to the situation with KRT1, HSPA5 was flagged up as a hit using the gel slice method because of absolute values. The absolute difference in detected spectra was exactly the same between the two methods (21 for UNR, 14 for IgG with the gel slice method and 10 for UNR and 3 for IgG with the on-bead method). In terms of fold change, a better value was obtained with the on-bead method even though HSPA5 was not detected as significant by that method.

The real problem here lies in the fact that detecting significance is a means to an end as opposed to an end in itself. Whilst it is impossible to know which, if either, of the two methods yields a true result when the true result is unknown and the two methods disagree, it is clear that both methods here can detect HSPA5.

Assuming that the UNR and IgG slices cover exactly the same range of protein masses, both the UNR and IgG slices will contain some shared non-specific proteins (which may also be UNR-specific proteins but are not bound in the classical antibody-bait-interactor manner). The IgG slice will contain a little more non-specific protein due to the space on the antibody that would otherwise bind UNR is free to bind to other things. The UNR slice will additionally contain some true UNR-interacting proteins. The relative proportions of specific UNR-bound proteins in a particular UNR gel slice will ultimately affect the ratio of specific and non-specific peptides

available to be fed into the mass spectrometer at a given point in time. That will, in turn, affect the ratio of the protein detected by UNR compared to IgG.

In conclusion, it was decided that there was not much difference between the two methods in terms of confidently predicting which proteins are true hits. It was ultimately decided that the benefits of cross-referencing putative interacting proteins by their relative molecular masses were more than compensated by the less biased quantification offered by the on-bead method coupled with the financial savings that could be used for additional experiments later in the project.

4.3 Main RIP-mass spectrometry experiments using HeLa cell lysates

Three sets of RIPs were carried out using HeLa cell lysates (arsenite and mock treated) prepared as outlined above and processed for mass spectrometry by the on-bead method. Initial trial injections of 2 µl for 30 minutes were followed by injections of 8 µl for 120 minutes. 15 mg of total protein were used per RIP with 12 µg of immunoprecipitating immunoglobulin (anti-UNR or control IgG) and 50 µl of Protein A slurry. The reactions were carried out in 15 ml tubes. As 20% of the RIP samples were set aside for mass spec and were ultimately resuspended in 50 µl, the 2 µl injections came from 120 µg of starting material, 96 ng of antibody and 400 nl of bead slurry. Likewise, the 8 µl injections came from 480 µg of starting material, 384 ng of antibody and 1.6 µl of bead slurry.

4.3.1 Data analysed using Scaffold software

In order to obtain an overview of the HeLa IP data that was obtained by mass spectrometry, it was decided to use the same software that was used with the trial data. Whilst it had been decided to use different software to quantify the data and generate ultimate observations and hypotheses based upon them, it was considered prudent to first check that the data were reproducible using similar tools to those used with the trial runs. To that end, the data were initially inputted into Scaffold (version Scaffold_4.5.3, Proteome Software Inc., Portland, OR, USA), as per the following Scaffold-generated report:

"DATABASE SEARCHING-- Tandem mass spectra were extracted by Unspecified version Unspecified. Charge state deconvolution and deisotoping were not performed. All MS/MS samples were analysed using Mascot (Matrix Science, London, England; version 2.5.0). Mascot was set up to search the human_uniprot_18June2015_Angela database (selected for Homo sapiens, unknown version, 90612 entries) assuming the digestion enzyme trypsin. Mascot was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 10.0 PPM. Carbamidomethyl of cysteine was specified in Mascot as a fixed modification. Oxidation of methionine was specified in Mascot as a variable modification.

"CRITERIA FOR PROTEIN IDENTIFICATION-- Scaffold (version Scaffold_4.5.3, Proteome Software Inc., Portland, OR, USA) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii *et al.*, 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters."

The raw protein data were quantified by Scaffold using its normalised total ion current setting. According to the Proteome Software website:

"The total ion current (TIC) is the sum of the areas under all the peaks contained in a MS/MS spectrum. Scaffold assumes that the area under a peak is proportional to the height of the peak and approximates the TIC value by summing the intensity of the peaks contained in the peak list associated to a MS/MS sample."

(https://proteome-software.wikispaces.com/Calculations+in+Scaffold+FAQs)

This assumption is then used to approximate the TIC value for a peptide by adding together the heights of all peaks in an MS/MS sample. It then sums all TIC values associated with peptides that were assigned to given proteins to give a value for at the protein or cluster level.

4.3.2 Data exported from Scaffold to Excel for further analysis

The resultant estimates of protein levels in each sample were then exported to Microsoft Excel (Microsoft Corp.). Following the removal of the trial injection data, paired two tail t-tests were carried out on the plus arsenite and minus arsenite samples separately. There were 316 protein clusters, of which 95 had no values recorded with the minus arsenite samples and 4 that had no values recorded in the plus arsenite samples. In order to get a general feel for the data, a liberal significance cut off point of p=0.1 was selected without a multiple testing correction. As some proteins were present at higher levels in the IgG samples, these were removed from further consideration. In addition to this, all proteins that were present in every UNR sample but absent in every IgG sample were considered, even if they were not considered significant by t-test (protein name shaded yellow). The proteins/protein clusters considered significant, by the parameters laid out above, from the HeLa minus arsenite samples are presented in Table 4.2B by p-value. The top 10 putative UNR interacting proteins are tabulated by t-test p-value with the total ion current values for each replicate (Table 4.2A). Likewise, the top ten putative UNR interacting proteins from the HeLa plus arsenite samples are presented in Table 4.2C and all putative interactors, by the criteria outlined above, are presented in Table 4.2D.

Table 4.2A: Top ten hits HeLa minus arsenite, by t-test p-value

Protein	HeLa1	HeLa2	HeLa3	HeLa4	HeLa5	HeLa6	p-value
SSBP4	0	0	0	685800	555720	676130	0.0043
UNRIP	0	0	0	2243600	1895500	1477100	0.0137
HUWE1	0	0	0	386300	265730	425550	0.0175
NARR	0	0	0	1972000	1237800	1377100	0.0210
RPL23A	0	0	0	118830	67403	86181	0.0263
HSPD1	0	0	130270	40965	34231	150430	0.0352
LDB1	0	0	0	442690	308230	214140	0.0400
UNR	0	0	0	3420100	1988100	1163000	0.0799
RPS5	155490	45419	0	273250	228700	56890	0.0822
SQSTM1	0	38981	0	181450	443410	160790	0.0858

N.B. IgG samples shaded lilac, repeat 1-3 running left to right (HeLa1-HeLa3). The corresponding UNR samples are shaded green and also run left to right (HeLa4-HeLa6). The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only. The full names for each protein can be found in Table 4.2B.

Table 4.2B: Putative UNR-interacting proteins from unstressed HeLa cells, by t-test p-value

Protein	p-value
Cluster of U3KPY3_HUMAN Single-stranded DNA-binding protein 4 (Fragment) GN=SSBP4	0.0043
Cluster of STRAP_HUMAN Serine-threonine kinase receptor-associated protein GN=STRAP	0.0137
Cluster of HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 GN=HUWE1	0.0175
NARR_HUMAN Ras-related protein RAB34, isoform NARR GN=RAB34	0.0210
Cluster of K7EJV9_HUMAN 60S ribosomal protein L23a (Fragment) GN=RPL23A	0.0263
CH60_HUMAN 60 kDa heat shock protein, mitochondrial GN=HSPD1	0.0352
Cluster of LDB1_HUMAN Isoform 2 of LIM domain-binding protein 1 GN=LDB1	0.0400
Cluster of CSDE1_HUMAN Isoform 4 of Cold shock domain-containing protein E1 GN=CSDE1	0.0799
M0R0F0_HUMAN 40S ribosomal protein S5 (Fragment) GN=RPS5	0.0822
Cluster of SQSTM_HUMAN Sequestosome-1 GN=SQSTM1	0.0858
LMO4_HUMAN LIM domain transcription factor LMO4 GN=LMO4	0.0872
Cluster of TFG_HUMAN Protein TFG GN=TFG	0.1878
Cluster of MYCB2_HUMAN E3 ubiquitin-protein ligase MYCBP2 GN=MYCBP2	0.2842

N.B. The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only.

Table 4.2C: Top ten hits HeLa plus arsenite, by t-test p-value

	HeLa7	HeLa8	HeLa9	HeLa10	HeLa11	HeLa12	p-value
DYNC1H1	0	0	2597900	778280	905160	3459800	0.0019
RPS2	121020	249370	111950	536230	704780	599190	0.0021
SSBP4	0	0	0	558350	425290	503750	0.0060
RPS3A	0	0	0	131180	172260	146250	0.0063
RPS4X	491630	569950	478130	1157600	1081000	1124200	0.0064
PRKDC	0	685150	6091400	797870	1753200	7197100	0.0094
HUWE1	0	0	0	667350	555970	458200	0.0114
LDB1	0	0	0	311660	268440	206310	0.0133
RPL28	0	0	175820	105510	160420	314300	0.0137
FAU	24481	23947	0	113530	101430	116240	0.0145

N.B. IgG samples shaded lilac, repeat 1-3 running left to right (HeLa7-HeLa9). The corresponding UNR samples are shaded green and also run left to right (HeLa10-HeLa12). The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only. The full names for each protein can be found in Table 4.2D.

Table 4.2D: Putative UNR-interacting proteins from arsenite stressed HeLa cells, by t-test p-value

Protein	p-value
Cluster of DYHC1_HUMAN Cytoplasmic dynein 1 heavy chain 1 GN=DYNC1H1	0.0019
Cluster of RS2_HUMAN 40S ribosomal protein S2 GN=RPS2	0.0021
Cluster of U3KPY3_HUMAN Single-stranded DNA-binding protein 4 (Fragment) GN=SSBP4	0.0060
D6RG13_HUMAN 40S ribosomal protein S3a (Fragment) GN=RPS3A	0.0063
Cluster of RS4X_HUMAN 40S ribosomal protein S4, X isoform GN=RPS4X	0.0064
PRKDC_HUMAN DNA-dependent protein kinase catalytic subunit GN=PRKDC	0.0094
Cluster of HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 GN=HUWE1	0.0114
Cluster of LDB1_HUMAN Isoform 2 of LIM domain-binding protein 1 GN=LDB1	0.0133
Cluster of H0YKD8_HUMAN 60S ribosomal protein L28 GN=RPL28	0.0137
RS30_HUMAN 40S ribosomal protein S30 GN=FAU	0.0145
A9Z1X7_HUMAN Serine/arginine repetitive matrix protein 1 GN=SRRM1	0.0161
GCN1L_HUMAN Translational activator GCN1 GN=GCN1L1	0.0211
Cluster of HSPB1_HUMAN Heat shock protein beta-1 GN=HSPB1	0.0234
Cluster of H0YI37_HUMAN ATP synthase subunit beta, mitochondrial (Fragment) GN=ATP5B	0.0323
Cluster of M0R2B0_HUMAN rRNA 2'-O-methyltransferase fibrillarin (Fragment) GN=FBL	0.0363
Cluster of CSDE1_HUMAN Isoform 4 of Cold shock domain-containing protein E1 GN=CSDE1	0.0369
Cluster of TADBP_HUMAN TAR DNA-binding protein 43 GN=TARDBP	0.0370
Cluster of RL27_HUMAN 60S ribosomal protein L27 GN=RPL27	0.0377
Cluster of RBM39_HUMAN RNA-binding protein 39 GN=RBM39	0.0379
Cluster of COF1_HUMAN Cofilin-1 GN=CFL1	0.0412
Cluster of SYYC_HUMAN Tyrosine--tRNA ligase, cytoplasmic GN=YARS	0.0445
E2AK2_HUMAN Interferon-induced, double-stranded RNA-activated protein kinase GN=EIF2AK2	0.0448
FAS_HUMAN Fatty acid synthase GN=FASN	0.0449
Cluster of TFG_HUMAN Protein TFG GN=TFG	0.0468
Cluster of STRAP_HUMAN Serine-threonine kinase receptor-associated protein GN=STRAP (Q9Y3F4)	0.0478
LARP4_HUMAN Isoform 5 of La-related protein 4 GN=LARP4	0.0482
Cluster of SYAC_HUMAN Alanine--tRNA ligase, cytoplasmic GN=AARS	0.0485
LMO4_HUMAN LIM domain transcription factor LMO4 GN=LMO4	0.0519
Cluster of HNRPL_HUMAN Heterogeneous nuclear ribonucleoprotein L GN=HNRNPL	0.0532
RS10_HUMAN 40S ribosomal protein S10 GN=RPS10	0.0572
NARR_HUMAN Ras-related protein RAB34, isoform NARR GN=RAB34	0.0578
FSCN1_HUMAN Fascin GN=FSCN1	0.0716
KEAP1_HUMAN Kelch-like ECH-associated protein 1 GN=KEAP1	0.0718

Cluster of H0YDD8_HUMAN 60S acidic ribosomal protein P2 (Fragment) GN=RPLP2	0.0732
Cluster of MCM7_HUMAN DNA replication licensing factor MCM7 GN=MCM7	0.0736
RL27A_HUMAN 60S ribosomal protein L27a GN=RPL27A	0.0759
Cluster of HNRPF_HUMAN Heterogeneous nuclear ribonucleoprotein F GN=HNRNPF	0.0774
Cluster of HNRPK_HUMAN Heterogeneous nuclear ribonucleoprotein K GN=HNRNPK	0.0782
Cluster of PFKAM_HUMAN ATP-dependent 6-phosphofructokinase, muscle type GN=PFKM	0.0785
RL7A_HUMAN 60S ribosomal protein L7a GN=RPL7A	0.0814
Cluster of ECM29_HUMAN Proteasome-associated protein ECM29 homolog GN=ECM29	0.0827
Cluster of D3DQV9_HUMAN Eukaryotic translation initiation factor 4 gamma 2 (Fragment) GN=EIF4G2	0.0853
KPRP_HUMAN Keratinocyte proline-rich protein GN=KPRP	0.0862
SND1_HUMAN Staphylococcal nuclease domain-containing protein 1 GN=SND1	0.0897
Cluster of ZCCHV_HUMAN Isoform 3 of Zinc finger CCCH-type antiviral protein 1 GN=ZC3HAV1	0.0947
Cluster of F5H5D3_HUMAN Tubulin alpha-1C chain GN=TUBA1C	0.0973
Cluster of HNRPQ_HUMAN Isoform 3 of Heterogeneous nuclear ribonucleoprotein Q GN=SYNCRIP	0.0978
JAK1_HUMAN Tyrosine-protein kinase JAK1 GN=JAK1	0.1113
Cluster of B4DUR8_HUMAN T-complex protein 1 subunit gamma GN=CCT3	0.1129
Cluster of SRRM2_HUMAN Serine/arginine repetitive matrix protein 2 GN=SRRM2	0.1218
Cluster of ZNFX1_HUMAN NFX1-type zinc finger-containing protein 1 GN=ZNFX1	0.1499
DHX9_HUMAN ATP-dependent RNA helicase A GN=DHX9	0.1631
Cluster of SQSTM_HUMAN Sequestosome-1 GN=SQSTM1	0.1741
Cluster of MCTP2_HUMAN Multiple C2 and transmembrane domain-containing protein 2 GN=MCTP2	0.1799
Cluster of H0YBN4_HUMAN Polyadenylate-binding protein 1 (Fragment) GN=PABPC1	0.1889
Cluster of FXR2_HUMAN Fragile X mental retardation syndrome-related protein 2 GN=FXR2	0.2260
Cluster of RBP2_HUMAN E3 SUMO-protein ligase RanBP2 GN=RANBP2	0.3235

N.B. The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only.

4.3.3 General consideration of the Excel analysis

Some things were immediately apparent from the analysis of this data set. Firstly, there appeared to be more putative UNR interactors in arsenite treated cells. Also, without accounting for hits obscured by differences in nomenclature between Mascot database labels, there were at least 5 putative hits from both the gel slice and on-bead HeLa trial runs that were considered significant in the larger scale HeLa minus arsenite experiment, by the criteria laid out in the text above (Tables 4.1, 4.2B). Those proteins were:

UNR, LDB1, LMO4, NARR and SQSTM1.

It was further noted that UNRIP was significant (t-test p-value <0.05) in both the plus and minus arsenite samples. UNRIP had been flagged up by both the gel slice and on-bead trial runs but a difference in the associated Accession number made it fail to appear on the list of hits for both as the hits had been compared by Accession number. That means that there was at least one more protein that was considered significant in at least three independent experiments, one with three biological repeats. Another point that was noted was that UNR was not found to associate significantly with SSBP3 in this experiment. The lowest p-value from the minus arsenite t-tests and the third lowest from the plus arsenite t-tests, however, was associated with the closely related SSBP4 protein. It has been noted, and will be discussed in more length later, that LDB1 and a number of the SSBP proteins, as well as other potential UNR-interactors such as LMO4, have been linked to a number of malignancies (see below). It was also noted that the five proteins (plus UNRIP) that were flagged up as putative UNR-interactors in the three different unstressed HeLa experiments were also suggested UNR interactors in the HeLa plus arsenite experiment (Tables 4.1, 4.2D).

4.3.4 Consideration of the minus arsenite data

It was noted that there was cross-over between the putative UNR interactors suggested by three different experiments using unstressed HeLa cell lysates – the gel slice trial run, the on-bead trial run and the main experiment with repeats. It was then decided to get a rough idea of how likely it would be that such a cross over could be obtained by chance alone.

In order to estimate this, it was first assumed that the joint hits from the trial run were correct and the probability of selecting the cross-over by chance was then estimated. It should be noted that this over-estimates any p-values as it assumes that the cross over between the trial runs has a probability of 1 when such an assumption is clearly invalid.

UNR was first removed from the list in Table 4.1. That was because, whilst it was reassuring to see UNR in all the UNR pull down samples, seeing it could not be a random event as the experiment was designed with the specific intention of pulling it down. UNRIP was then added to the list for the reasons stated above, making 17 proteins. This was then compared to the 12 non-UNR proteins in Table 4.2B that also contained putative UNR-interacting proteins from non-arsenite-treated HeLa cells.

It was noted that 5 proteins were suggested by both the trial experiments and the main HeLa minus arsenite experiment (LDB1, LMO4, NARR, SQSTM1 and UNRIP). The Scaffold report for the trial run proteins stated that the Uniprot database used (the earlier of the two, implying fewer proteins) contained 134431 protein entries. As many of those will be repeated, a more conservative estimate of 20000 proteins encoded by the human genome was considered. The estimate was then framed as a question in the format of “given that there are 20000 balls in a bag, of which 17 are red, and 12 balls are drawn at random, what is the probability that exactly 5 drawn balls are red?”.

Taking the 17 proteins suggested by the trial runs and comparing this to the 12 found by the larger scale experiment, this implies that the probability of seeing 5 on both lists by chance alone

is the product of the number of unique combinations of 5 red balls chosen from 12 (12!/(5!x(12-5)!)) and the probability of drawing any 5 red balls among a random selection of 12 in 20000:

$$p=\left(\frac{12x11x10x9x8}{5x4x3x2x1}\right)x\left(\frac{(17x16x15x14x13)x(19983x19982x...x19977)}{20000x19999x...x19989}\right)$$

This is around $1.83*10^{-13}$, so it seems that there is a genuine signal present and that the techniques we used were likely to be discovering some real UNR interacting proteins. Another way of generating the exact probability is to use the hypergeometric test. Using the statistical language R (https://www.r-project.org/about.html), two hypergeometric tests were carried out. The first was for choosing more than 4 red balls from a subset of 17 red balls within a total population of 20000 balls with 12 random selections. The second was the same calculation for choosing more than 5 red balls. The second value was taken away from the first, thereby giving the probability of drawing exactly 5 red balls.

```
>phyper(4,17,20000-17,12,lower.tail=FALSE)-phyper(5,17,20000-17,12,lower.tail=FALSE)
[1] 1.831043e-13
```

It is customary to give the probability of observing a value at least as extreme as a given observation as opposed to that of observing an exact value. This is obtained by not subtracting the second hypergeometric test value. Due to the high total population of proteins (balls), the probability of seeing more extreme values falls off quickly and, as a result, the probability of seeing an observation as least as extreme as 5 rounds to the same value to 3 significant figures:

```
> phyper(4,17,20000-17,12,lower.tail=FALSE)
[1] 1.832327e-13
```

One potential criticism of using an approach such as this surrounds the assumption that all 20000 proteins are equally likely to be expressed and detected. It could be that a much smaller selection of proteins could be detected. To see the effect of population size on the probability

of seeing a degree of cross-over at least as extreme as was observed, the test was repeated using total populations of 10000, 5000, 1000 and 500 proteins:

```
> phyper(4,17,10000-17,12,lower.tail=FALSE)
[1] 5.845847e-12
> phyper(4,17,5000-17,12,lower.tail=FALSE)
[1] 1.859436e-10
> phyper(4,17,1000-17,12,lower.tail=FALSE)
[1] 5.534081e-07
> phyper(4,17,500-17,12,lower.tail=FALSE)
[1] 1.663646e-05
```

Assuming that there were only 500 proteins in the HeLa minus arsenite samples that could be detected by the mass spectrometer, the probability of choosing more than 4 red balls from a subset of 17 red balls within a total population of 500 balls with 12 random selections would still be significant ($p = 1.66 \times 10^{-5}$). This strongly suggests that the cross-over was not random and that the 5 proteins were likely to be true UNR-interactors.

4.3.5 Consideration of the plus arsenite data

The top hit by t-test p-value for arsenite treated HeLa cells was dynein heavy chain 1. This was considered interesting as DHC1 had been shown, together with microtubules, to be required for the proper establishment of stress granules in HeLa and other cell lines. (Loschi *et al.*, 2009). Another promising sign was that there was a degree of cross-over between the plus and minus arsenite samples with ten of the top 25 putative hits for the minus arsenite samples also being top putative hits for the plus arsenite samples and that these included the five hits previously noted to be constant between the trial runs and the main minus arsenite experiments (LDB1, LMO4, NARR, SQSTM1 and UNRIP). (Table 4.3).

Table 4.3: Putative UNR interacting proteins in both plus and minus arsenite HeLa samples

Cluster of U3KPY3_HUMAN Single-stranded DNA-binding protein 4 (Fragment) (GN=SSBP4)
Cluster of CSDE1_HUMAN Isoform 4 of Cold shock domain-containing protein E1 GN=CSDE1
Cluster of HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 (GN=HUWE1)
Cluster of LDB1_HUMAN Isoform 2 of LIM domain-binding protein 1 (GN=LDB1)
Cluster of SQSTM_HUMAN Sequestosome-1 (GN=SQSTM1)
Cluster of STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)
Cluster of TFG_HUMAN Protein TFG (GN=TFG)
LMO4_HUMAN LIM domain transcription factor LMO4 (GN=LMO4)
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)
PRKDC_HUMAN DNA-dependent protein kinase catalytic subunit (GN=PRKDC)

4.3.6 Consideration of some of the most reproducible putative hits

UNR and UNRIP have been discussed previously in the introduction. The DNA that encodes NARR, standing for 'nine amino-acid residue-repeats', overlaps that of the *RAB34* proto-oncogene (Zougman *et al.*, 2011). It is a nucleolar protein that associates with rDNA clusters and is well conserved among the mammals (Zougman *et al.*, 2011). The thirteen nine amino-acid repeats of the human protein contain one serine or threonine residue that is phosphorylated during M-phase (Zougman *et al.*, 2011). Should UNR be interacting with a nucleolar protein, one of three things must be true:

1) UNR is interacting with the protein inside the nucleus
2) UNR interacts with the protein in the cytoplasm prior to it entering the nucleus
3) UNR interacts with the protein during mitosis when nuclear proteins can come into contact with cytoplasmic proteins.

Whilst it is not clear which of these is true, it has been noted elsewhere that UNR concentration can spike during mitosis (Schepens *et al.*, 2007). It should also be noted that the production of cell lysate can rupture intracellular membranes, thereby creating a non-physiological environment in which nuclear and non-nuclear proteins can come into contact with each other.

Sequestosome 1 is a well conserved metazoan stress-induced protein involved in regulating a variety of cellular functions including selective autophagy, lipogenesis and apoptosis as well as a number of signalling pathways including activation of both the Raptor-containing mTOR complex 1 and NF-κB. It also contains a zinc finger and can interact with DNA (reviewed in Katsuragi et al. 2015). Interestingly, it has also been reported to be involved in the development of Alzheimer's disease (Salminen *et al.*, 2012) and tumour proliferation and metastasis through an interaction with, and stabilisation of, the EMT-promoting Twist1 (Qiang *et al.*, 2014). Many of these processes, such as apoptosis, metastasis and Alzheimer's disease have also been associated with UNR.

LIM domain binding protein 1 (LDB1) has a wide range of cellular functions, of which only a few will be mentioned here for the sake of brevity. Ldb1 is associated with pancreatic development and the expression of certain genes, such as insulin, in conjunction with the transcription factor Islets 1 (Isl1) and other genes, such as Glucose transporter type 2 (Glut2) in an Isl1-independent manner (Hunter *et al.*, 2013). UNR has also been associated with glucose levels and diabetes (Xavier *et al.*, 2014). Ldb1 is required for erythropoiesis (Li *et al.*, 2010). Finally, LDB1 has been shown to co-immunoprecipitate with Tripartite Motif-Containing Protein 33 (TRIM33) (Bai *et al.*, 2010). Furthermore, TRIM33 has been shown to decrease the cellular LDB1 level (Howard *et al.*, 2010). This is interesting on several UNR-related levels:

1) UNR and TRIM33 lie in extremely close proximity on human chromosome 1.
2) UNR and TRIM33 have been functionally linked to autism (Xia *et al.*, 2013).
3) Like UNR, TRIM33 is related to embryonic development (see below).
4) Like UNRIP (see above), TRIM33 is linked TGF-β signalling generally (see below).
5) Like UNR, TRIM33 is linked to erythropoiesis (see below) and to Diamond Blackfan Anaemia in particular (Ge *et al.*, 2015).

TRIM33 interacts with SMAD proteins in various biological settings, including development and haematopoiesis ((Ransom *et al.*, 2004)Dupont et al. 2005; He et al. 2006; Morsut et al. 2010). TRIM33 inhibits signalling through SMAD4 via a ubiquitylation event that reduces the proportion of SMAD4 found in the nucleus, both in the presence and absence of TGF-β1 (Dupont *et al.*, 2005). It also further shown that TRIM33 monoubiquitylation of SMAD4 destabilises the interaction between SMAD4 and activated SMAD2/3, and that the ubiquitylation is antagonised by USP9x (Dupont *et al.*, 2009).

LIM domain only 4 is a transcription factor that has been reported to be involved in a variety of cancers. For example, an LDB1-LMO4 interaction was suggested to be involved in oral cavity squamous carcinoma (Mizunuma *et al.*, 2003). Likewise, LMO4 was shown to be frequently upregulated in invasive breast carcinomas (Visvader *et al.*, 2001) and to interact with and negatively regulate the tumour suppressor BRCA1 protein (Sum *et al.*, 2002). At least one LMO4 and LDB1-containing complex, that also contains GATA6, functions as a tumour suppressor by activating p21 in a potentially TP53-independent manner (Setogawa *et al.*, 2006). As this involves GATA6, it thereby provides an indirect link to the work of Elatmani and others in the Jacquemin-Sablon lab (Elatmani *et al.*, 2011).

Ssbp3 stabilises the interaction between Ldb1 and Lhx1 in early development (Costello *et al.*, 2015). Also, both this protein and SSBP4 were both shown to co-immunoprecipitate with TRIM33 (Bai *et al.*, 2010).

4.4 Choice of additional cell types to extend the experiment

Having made some initial observations using HeLa cells, it was decided to extend the work to carry out repeats with other cell types. That was with the intention of discovering whether the observations were cell type specific.

The osteosarcoma cell line U2OS was selected because it was available in the lab and was known to express wild type TP53. That was important as some exciting findings in the lab had potentially linked UNR to TP53 (Ray, Ó Catnaigh and Anderson, manuscript in preparation). To get an insight into any potential functional interaction between UNR and TP53, it was also decided to carry

out RIPs using the related TP53-null osteosarcoma cell line, SaOS-2 (Diller *et al.*, 1990). It should also be considered that, while U2OS contains wild type TP53, it also over-expresses HDM2 (Florenes *et al.*, 1994).

It had been noted that the p-values associated with many putative hits observed to this point were too high to pass a multiple testing correction. It was therefore considered welcome that the same proteins were coming up in multiple independent experiments. It was hoped that seeing the same hits in additional experiments in other cell types would further confirm that the approach taken was valid and that many of the observed hits were unlikely to be false positives.

In short, the additional work was undertaken in the hope of lending further support to findings already made in HeLa cells as well as finding putative UNR interactors in other cell types.

4.5 Main RIP-mass spectrometry experiments using U2OS cell lysates

Three sets of RIPs were carried out using U2OS cell lysates (arsenite and mock treated) prepared as outlined above and processed for mass spectrometry by the on-bead method. Initial trial injections of 2 µl for 30 minutes were followed by injections of 8 µl for 120 minutes. 7 mg of total protein was used per RIP with 12 µg of immunoprecipitating immunoglobulin (anti-UNR or control IgG) and 50 µl of Protein A slurry. The reactions were carried out in 15 ml tubes. As 20% of the RIP samples were set aside for mass spec and were ultimately resuspended in 50 µl, the 2 µl injections came from 56 µg of starting material, 96 ng of antibody and 400 nl of bead slurry. Likewise, the 8 µl injections came from 224 µg of starting material, 384 ng of antibody and 1.6 µl of bead slurry.

A Scaffold file was made as before; the publication report for the U2OS run was exactly the same as for the HeLa report and the same pipeline was used for analysis as had been used for the HeLa samples.

There were 368 clusters, of which 115 were not detected in any of the minus arsenite samples and 8 were not detected in any of the plus arsenite samples. The data were sorted using the same significance criteria as for the HeLa samples (see section 4.3.2).

4.5.1 Putative UNR interactors found using U2OS cell lysates

There were 21 putative hits for the U2OS minus arsenite samples (Table 4.4B), the top ten of which by p-value are presented in Table 4.4A. There were 76 putative hits for the U2OS plus arsenite samples (Table 4.4D), of which the top ten by p-value are tabulate below (Table 4.4C). There were 9 proteins that were putative hits under both conditions (Table 4.4E).

Table 4.4A: Top ten hits U2OS minus arsenite, by t-test p-value

Protein	U2OS1	U2OS2	U2OS3	U2OS4	U2OS5	U2OS6	p-value
EWSR1	220690	79529	100240	316360	171140	180290	0.0027
VIM	102930	164410	127120	138130	192810	160180	0.0039
LDB1	0	0	0	245600	208200	300600	0.0112
RPS27	64800	111000	122000	104000	165000	190000	0.0234
UNR	0	0	0	644760	328380	580240	0.033
ACTN1	0	367530	0	85461	453590	45127	0.0334
DDX3X	591330	228460	573350	970920	411940	933630	0.0387
TMEM97	0	0	0	87815	59805	128300	0.0437
ATP5A1	151760	872820	76693	1127100	1344900	683320	0.045
RPS7	189240	0	0	224140	55669	79267	0.0476
HSPA5	0	449500	0	51703	578090	99651	0.0532

N.B. IgG samples shaded lilac repeat 1-3 running left to right (U2OS1-U2OS3). The corresponding UNR samples are shaded green and also run left to right (U2OS4-U2OS6). The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only.

Table 4.4B: Putative UNR-interacting proteins from unstressed U2OS cells, by t-test p-value

Protein	p-value
Cluster of B0QYK0_HUMAN RNA-binding protein EWS (GN=EWSR1)	0.0027
Cluster of VIME_HUMAN Vimentin (GN=VIM)	0.0039
Cluster of LDB1_HUMAN LIM domain-binding protein 1 (GN=LDB1)	0.0112
RS27_HUMAN 40S ribosomal protein S27 (GN=RPS27)	0.0234
CSDE1_HUMAN Isoform 4 of Cold shock domain-containing protein E1 (GN=CSDE1)	0.0330
Cluster of ACTN1_HUMAN Alpha-actinin-1 (GN=ACTN1)	0.0334
Cluster of A0A0D9SF53_HUMAN ATP-dependent RNA helicase DDX3X (GN=DDX3X)	0.0387
TMM97_HUMAN Transmembrane protein 97 (GN=TMEM97)	0.0437
ATPA_HUMAN ATP synthase subunit alpha, mitochondrial (GN=ATP5A1)	0.0450
RS7_HUMAN 40S ribosomal protein S7 (GN=RPS7)	0.0476
GRP78_HUMAN 78 kDa glucose-regulated protein (GN=HSPA5)	0.0532
Cluster of SSBP3_HUMAN Single-stranded DNA-binding protein 3 (GN=SSBP3)	0.0544
Cluster of HUWE1_HUMAN Isoform 3 of E3 ubiquitin-protein ligase HUWE1 (GN=HUWE1)	0.0686
RS8_HUMAN 40S ribosomal protein S8 (GN=RPS8)	0.0768
SRSF2_HUMAN Serine/arginine-rich splicing factor 2 (GN=SRSF2)	0.0812
RL29_HUMAN 60S ribosomal protein L29 (GN=RPL29)	0.0845
H4_HUMAN Histone H4 (GN=HIST1H4A)	0.0850
STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)	0.1160
E9PK91_HUMAN Bcl-2-associated transcription factor 1 (GN=BCLAF1)	0.1264
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)	0.1670
U2AF2_HUMAN Splicing factor U2AF 65 kDa subunit (GN=U2AF2)	0.1769

N.B. The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only.

Table 4.4C: Top ten hits U2OS plus arsenite, by t-test p-value

Protein	U2OS7	U2OS8	U2OS9	U2OS10	U2OS11	U2OS12	p-value
ANXA2	0	0	0	91468	90853	81731	0.0013
UNRIP	0	0	0	3243000	3005700	3508700	0.002
UNR	0	0	0	3774400	3703900	4439600	0.0035
HSPA9	0	0	0	161500	127070	186940	0.0118
RPL11	35120	59251	118000	102600	157590	193990	0.0128
TUBB6	0	85370	347680	626070	893540	1367000	0.0188
LDB1	0	0	0	285440	209560	356440	0.0216
RPS2	332580	172460	302870	733490	874400	880930	0.0235
NARR	0	0	0	282330	159030	290080	0.029
RBM14	0	0	0	105720	72114	138340	0.0314
TUBB	407100	1326300	1790500	4629400	3511500	4856900	0.0332

N.B. IgG samples shaded lilac repeat 1-3 running left to right (U2OS7-U2OS9). The corresponding UNR samples are shaded green and also run left to right (U2OS10-U2OS12). The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only.

Table 4.4D: Putative UNR-interacting proteins from arsenite stressed U2OS cells, by t-test p-value

Protein	p-value
Cluster of ANXA2_HUMAN Annexin A2 (GN=ANXA2)	0.0013
STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)	0.0020
CSDE1_HUMAN Isoform 4 of Cold shock domain-containing protein E1 (GN=CSDE1)	0.0035
Cluster of GRP75_HUMAN Stress-70 protein, mitochondrial (GN=HSPA9)	0.0118
RL11_HUMAN 60S ribosomal protein L11 (GN=RPL11)	0.0128
TBB6_HUMAN Tubulin beta-6 chain (GN=TUBB6)	0.0188
Cluster of LDB1_HUMAN LIM domain-binding protein 1 (GN=LDB1)	0.0216
Cluster of RS2_HUMAN 40S ribosomal protein S2 (GN=RPS2)	0.0235
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)	0.0290
Cluster of RBM14_HUMAN RNA-binding protein 14 (GN=RBM14)	0.0314
Cluster of TBB5_HUMAN Tubulin beta chain (GN=TUBB)	0.0332
Cluster of C9JJQ8_HUMAN Tubulin alpha-4A chain (Fragment) (GN=TUBA4A)	0.0344
A0A087WXM6_HUMAN 60S ribosomal protein L17 (Fragment) (GN=RPL17)	0.0348
Cluster of RINI_HUMAN Ribonuclease inhibitor (GN=RNH1)	0.0371
Cluster of HUWE1_HUMAN Isoform 3 of E3 ubiquitin-protein ligase HUWE1 (GN=HUWE1)	0.0400
RL8_HUMAN 60S ribosomal protein L8 (GN=RPL8)	0.0444
Cluster of HS71A_HUMAN Heat shock 70 kDa protein 1A (GN=HSPA1A)	0.0461
Cluster of B5MDF5_HUMAN GTP-binding nuclear protein Ran (GN=RAN)	0.0468
Cluster of J3KQ22_HUMAN Pyrroline-5-carboxylate reductase (Fragment) (GN=PYCR1)	0.0510
HSPB1_HUMAN Heat shock protein beta-1 (GN=HSPB1)	0.0526
Cluster of RS30_HUMAN 40S ribosomal protein S30 (GN=FAU)	0.0561
DYHC1_HUMAN Cytoplasmic dynein 1 heavy chain 1 (GN=DYNC1H1)	0.0563
Cluster of SSBP3_HUMAN Single-stranded DNA-binding protein 3 (GN=SSBP3)	0.0565
AHNK_HUMAN Neuroblast differentiation-associated protein AHNAK (GN=AHNAK)	0.0567
RS8_HUMAN 40S ribosomal protein S8 (GN=RPS8)	0.0580
MCM7_HUMAN DNA replication licensing factor MCM7 (GN=MCM7)	0.0634
Cluster of FLNC_HUMAN Filamin-C (GN=FLNC)	0.0650
Cluster of H2B1A_HUMAN Histone H2B type 1-A (GN=HIST1H2BA)	0.0655
Cluster of H0Y449_HUMAN Nuclease-sensitive element-binding protein 1 (Fragment) (GN=YBX1)	0.0681
Cluster of RS3_HUMAN 40S ribosomal protein S3 (GN=RPS3)	0.0682
Cluster of H7C1H2_HUMAN 26S proteasome non-ATPase regulatory subunit 2 (Fragment) (GN=PSMD2)	0.0686
Cluster of RS9_HUMAN 40S ribosomal protein S9 (GN=RPS9)	0.0704
Cluster of BAG2_HUMAN BAG family molecular chaperone regulator 2 (GN=BAG2)	0.0731
FAS_HUMAN Fatty acid synthase (GN=FASN)	0.0755

Cluster of HNRPK_HUMAN Heterogeneous nuclear ribonucleoprotein K (GN=HNRNPK)	0.0757
K7EJV9_HUMAN 60S ribosomal protein L23a (Fragment) (GN=RPL23A)	0.0811
Cluster of E9PFH4_HUMAN Transportin-3 (GN=TNPO3)	0.0824
Cluster of TFG_HUMAN Protein TFG (GN=TFG)	0.0835
ATPA_HUMAN ATP synthase subunit alpha, mitochondrial (GN=ATP5A1)	0.0848
MSH6_HUMAN DNA mismatch repair protein Msh6 (GN=MSH6)	0.0887
Cluster of PCBP1_HUMAN Poly(rC)-binding protein 1 (GN=PCBP1)	0.0887
RS14_HUMAN 40S ribosomal protein S14 (GN=RPS14)	0.0916
Cluster of ADT2_HUMAN ADP/ATP translocase 2 (GN=SLC25A5)	0.0986
EMD_HUMAN Emerin (GN=EMD)	0.1043
Cluster of A0A0A0MSG2_HUMAN Four and a half LIM domains protein 2 (GN=FHL2)	0.1119
Cluster of COF1_HUMAN Cofilin-1 (GN=CFL1)	0.1119
Cluster of PLEC_HUMAN Plectin (GN=PLEC)	0.1119
Cluster of SQSTM_HUMAN Sequestosome-1 (GN=SQSTM1)	0.1147
Cluster of TRIP6_HUMAN Thyroid receptor-interacting protein 6 (GN=TRIP6)	0.1148
Cluster of IPO4_HUMAN Importin-4 (GN=IPO4)	0.1188
SSRD_HUMAN Translocon-associated protein subunit delta (GN=SSR4)	0.1205
Cluster of TIF1B_HUMAN Transcription intermediary factor 1-beta (GN=TRIM28)	0.1224
U2AF2_HUMAN Splicing factor U2AF 65 kDa subunit (GN=U2AF2)	0.1312
Cluster of ATPO_HUMAN ATP synthase subunit O, mitochondrial (GN=ATP5O)	0.1335
Cluster of MPCP_HUMAN Phosphate carrier protein, mitochondrial (GN=SLC25A3)	0.1414
Cluster of M0QXS5_HUMAN Heterogeneous nuclear ribonucleoprotein L (Fragment) (GN=HNRNPL)	0.1428
Cluster of J3QL54_HUMAN Nuclear pore complex protein Nup85 (GN=NUP85)	0.1526
Cluster of F8W6I7_HUMAN Heterogeneous nuclear ribonucleoprotein A1 (GN=HNRNPA1)	0.1541
Cluster of H3BVG0_HUMAN Nuclear pore complex protein Nup93 (GN=NUP93)	0.1559
Cluster of Q5JR05_HUMAN Rho-related GTP-binding protein RhoC (GN=RHOC)	0.1593
Cluster of ATD3A_HUMAN Isoform 2 of ATPase family AAA domain-containing protein 3A (GN=ATAD3A)	0.1635
EF2_HUMAN Elongation factor 2 (GN=EEF2)	0.1708
Cluster of RL28_HUMAN 60S ribosomal protein L28 (GN=RPL28)	0.1798
Cluster of E9PR16_HUMAN Nuclear pore complex protein Nup160 (Fragment) (GN=NUP160)	0.1936
Cluster of 1433B_HUMAN 14-3-3 protein beta/alpha (GN=YWHAB)	0.1958
Cluster of IMB1_HUMAN Importin subunit beta-1 (GN=KPNB1)	0.2152
Cluster of H7C2U6_HUMAN Protein NipSnap homolog 1 (Fragment) (GN=NIPSNAP1)	0.2295
Cluster of C9JZR2_HUMAN Catenin delta-1 (GN=CTNND1)	0.2330
Cluster of B4DR61_HUMAN Protein transport protein Sec61 subunit alpha isoform 1 (GN=SEC61A1)	0.2499
Cluster of RBP2_HUMAN E3 SUMO-protein ligase RanBP2 (GN=RANBP2)	0.2538
RL4_HUMAN 60S ribosomal protein L4 (GN=RPL4)	0.2648

Cluster of PLAP_HUMAN Phospholipase A-2-activating protein (GN=PLAA)	0.2931
A0A0A0MSX9_HUMAN Isoleucine--tRNA ligase, cytoplasmic (GN=IARS)	0.3092
Cluster of Q5SZU1_HUMAN D-3-phosphoglycerate dehydrogenase (GN=PHGDH)	0.3112
SYEP_HUMAN Bifunctional glutamate/proline--tRNA ligase (GN=EPRS)	0.3125
MVP_HUMAN Major vault protein (GN=MVP)	0.3478

N.B. The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only.

Table 4.4E: Putative UNR interacting proteins in both plus and minus arsenite U2OS samples

STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)
CSDE1_HUMAN Isoform 4 of Cold shock domain-containing protein E1 (GN=CSDE1)
Cluster of LDB1_HUMAN LIM domain-binding protein 1 (GN=LDB1)
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)
Cluster of HUWE1_HUMAN Isoform 3 of E3 ubiquitin-protein ligase HUWE1 (GN=HUWE1)
Cluster of SSBP3_HUMAN Single-stranded DNA-binding protein 3 (GN=SSBP3)
RS8_HUMAN 40S ribosomal protein S8 (GN=RPS8)
ATPA_HUMAN ATP synthase subunit alpha, mitochondrial (GN=ATP5A1)
U2AF2_HUMAN Splicing factor U2AF 65 kDa subunit (GN=U2AF2)

At first sight, there appeared to be several recurrent proteins that could be true UNR-interactors. UNRIP, LDB1, NARR and HUWE1 were considered putative hits from both the plus and minus arsenite U2OS samples (Table 4.4E) and had also been putative hits from the plus and minus arsenite HeLa samples (Table 4.3). On top of this, it was noted that SSBP3 was a putative UNR-interactor in stressed and unstressed U2OS cells (Table 4.4E). This had also been a hit in both of the trial runs that used unstressed HeLa cells (Table 4.1).

Among other observations, it was noted that 78 kDa glucose-regulated protein (HSPA5) was a hit for the minus arsenite samples, as it was for the HeLa gel slice trial run (Figure 4.12). Also, SQSTM1 was present in the list of putative hits for U2OS plus arsenite, as was DYNC1H1, the top hit for HeLa plus arsenite. See the Progenesis/quantitative analysis section below for further consideration of the similarities and differences between the cell types.

Having had an initial look at the U2OS data, it was decided to have a similar look at the SaOS-2 data prior to moving on to analyse the main data further using Progenesis.

4.6 Main RIP-mass spectrometry experiments using SaOS-2 cell lysates

Three sets of RIPs were carried out using SaOS-2 cell lysates (arsenite and mock treated) prepared as outlined above and processed for mass spectrometry by the on-bead method. Initial trial injections were not carried out on these sample. The only run carried out was the main run with injections of 8 µl for 120 minutes. 6 mg of total protein was used per RIP with 12 µg of immunoprecipitating immunoglobulin (anti-UNR or control IgG) and 50 µl of Protein A slurry. The reactions were carried out in 15 ml tubes. As 20% of the RIP samples were set aside for mass spec and were ultimately resuspended in 50 µl, the 8 µl injections came from 192 µg of starting material, 384 ng of antibody and 1.6 µl of bead slurry.

Again, the Scaffold publication report was exactly the same for the SaOS-2 run as for the previous two cell types and the same steps were used in analysing the data as with the other cell types. There were 399 clusters, 106 of which were not found in the minus arsenite samples and 1 was not present in the plus arsenite data.

4.6.1 Putative UNR interactors found using SaOS-2 cell lysates

There were 41 putative hits for the SaOS-2 minus arsenite samples (Table 4.5B), of which the top ten by p-value are presented below (Table 4.5A). There were 146 putative hits for the SaOS-2 plus arsenite samples (Table 4.5D), of which the top ten by p-value are documented below (Table 4.5C). As there were a larger number of hits than before, it was decided to present only gene names with p-values to save table space (Tables 4.5B and 4.5D). There were 33 proteins that were putative hits under both conditions (Table 4.5E).

Table 4.5A: Top ten hits SaOS-2 minus arsenite, by t-test p-value

	SaOS1	SaOS2	SaOS3	SaOS4	SaOS5	SaOS6	p-value
RPL27	0	0	0	117740	135440	123520	0.0017
LMO4	0	0	0	180810	181340	158960	0.0018
DDX41	0	0	118210	114270	110680	255840	0.0049
SSBP3	0	0	0	632010	526810	726910	0.0083
LDB1	0	0	0	701290	809560	986080	0.0098
PGAM5	0	0	0	342290	553430	441620	0.0182
SQSTM1	0	0	0	358270	476870	609300	0.0219
ARHGEF2	0	0	0	210840	207810	334010	0.0264
RPS27	0	0	0	31990	62522	43473	0.0355
HIST1H4A	83542	0	0	287660	351420	183040	0.0433
HUWE1	0	0	0	3476800	2617600	1487100	0.0482

N.B. IgG samples shaded lilac, repeat 1-3 running left to right (1SaOS-3SaOS). The corresponding UNR samples are shaded green and also run left to right (4SaOS-6SaOS). The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only.

Table 4.5B: Putative UNR-interacting proteins from unstressed SaOS-2 cells, by t-test p-value

	p-value
RL27_HUMAN 60S ribosomal protein L27 (GN=RPL27)	0.0017
LMO4_HUMAN LIM domain transcription factor LMO4 (GN=LMO4)	0.0018
Cluster of J3KNN5_HUMAN Probable ATP-dependent RNA helicase DDX41 (Fragment) (GN=DDX41)	0.0049
Cluster of SSBP3_HUMAN Single-stranded DNA-binding protein 3 (GN=SSBP3)	0.0083
LDB1_HUMAN Isoform 3 of LIM domain-binding protein 1 (GN=LDB1)	0.0098
PGAM5_HUMAN Serine/threonine-protein phosphatase PGAM5, mitochondrial (GN=PGAM5)	0.0182
Cluster of SQSTM_HUMAN Sequestosome-1 (GN=SQSTM1)	0.0219
Cluster of ARHG2_HUMAN Rho guanine nucleotide exchange factor 2 (GN=ARHGEF2)	0.0264
RS27_HUMAN 40S ribosomal protein S27 (GN=RPS27)	0.0355
H4_HUMAN Histone H4 (GN=HIST1H4A)	0.0433
Cluster of HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 (GN=HUWE1)	0.0482
RS18_HUMAN 40S ribosomal protein S18 (GN=RPS18)	0.0510
Cluster of SRSF6_HUMAN Serine/arginine-rich splicing factor 6 (GN=SRSF6)	0.0550
RS14_HUMAN 40S ribosomal protein S14 (GN=RPS14)	0.0590
Cluster of F1T0I1_HUMAN Protein transport protein Sec16A (GN=SEC16A)	0.0606
EWS_HUMAN RNA-binding protein EWS (GN=EWSR1)	0.0621
Cluster of C9J0D1_HUMAN Histone H2A (GN=H2AFV)	0.0653
RL23_HUMAN 60S ribosomal protein L23 (GN=RPL23)	0.0665
Cluster of A9Z1X7_HUMAN Serine/arginine repetitive matrix protein 1 (GN=SRRM1)	0.0862
Cluster of RS3_HUMAN 40S ribosomal protein S3 (GN=RPS3)	0.0868
Cluster of J3KQE5_HUMAN GTP-binding nuclear protein Ran (Fragment) (GN=RAN)	0.0910
Cluster of F5H5D3_HUMAN Tubulin alpha-1C chain (GN=TUBA1C)	0.0988
Cluster of CO4A_HUMAN Complement C4-A (GN=C4A)	0.1061
Cluster of SRRM2_HUMAN Serine/arginine repetitive matrix protein 2 (GN=SRRM2)	0.1117
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)	0.1119
A6NNN6_HUMAN Pericentriolar material 1 protein (GN=PCM1)	0.1226
Cluster of RS6_HUMAN 40S ribosomal protein S6 (GN=RPS6)	0.1228
Cluster of M0R210_HUMAN 40S ribosomal protein S16 (GN=RPS16)	0.1230
Cluster of CSDE1_HUMAN Cold shock domain-containing protein E1 (GN=CSDE1)	0.1567
Cluster of TBB5_HUMAN Tubulin beta chain (GN=TUBB)	0.1609
Cluster of H2B1D_HUMAN Histone H2B type 1-D (GN=HIST1H2BD)	0.1640
Cluster of A0A075B7D9_HUMAN TATA-binding protein-associated factor 2N (GN=TAF15)	0.1674
Cluster of HSP7C_HUMAN Heat shock cognate 71 kDa protein (GN=HSPA8)	0.1701

RS4X_HUMAN 40S ribosomal protein S4, X isoform (GN=RPS4X)	0.1816
Cluster of J3KP15_HUMAN Serine/arginine-rich-splicing factor 2 (Fragment) (GN=SRSF2)	0.1857
SETMR_HUMAN Histone-lysine N-methyltransferase SETMAR (GN=SETMAR)	0.1926
GRP78_HUMAN 78 kDa glucose-regulated protein (GN=HSPA5)	0.1956
MCM7_HUMAN DNA replication licensing factor MCM7 (GN=MCM7)	0.2035
STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)	0.2175
P4HA2_HUMAN Prolyl 4-hydroxylase subunit alpha-2 (GN=P4HA2)	0.2547
MYCB2_HUMAN E3 ubiquitin-protein ligase MYCBP2 (GN=MYCBP2)	0.2954

N.B. The t-test p-values are for a two tailed paired t-test without multiple testing correction. Yellow shading denotes proteins that were recorded in UNR samples only.

Table 4.5C: Top ten hits SaOS-2 plus arsenite, by t-test p-value

	7SaOS	8SaOS	9SaOS	10SaOS	11SaOS	12SaOS	p-value
UNR	0	0	0	6458000	6436900	6551900	0.00003
HSPA8	0	0	0	2244300	2392900	2198100	0.0007
TUBA1C	49380	332020	337080	4016900	4695300	4387200	0.0009
IDH2	0	0	0	131460	150880	139760	0.0016
LDB1	0	0	0	1098800	850600	907790	0.0062
MMS19	0	0	0	119790	111960	90865	0.0064
RPL10	0	0	0	197570	153870	200000	0.0066
UNRIP	0	0	0	5408200	4004600	5010700	0.0075
FBL	0	0	0	180900	240000	241440	0.0081
PDXDC1	0	0	0	149730	203760	198450	0.0086
RAP1B	0	0	0	130290	183090	159960	0.0092

N.B. IgG samples shaded lilac, repeat 1-3 running left to right (7SaOS-9SaOS). The corresponding UNR samples are shaded green and also run left to right (10SaOS-12SaOS). The t-test p-values are for a two tailed paired t-test without multiple testing correction.

Table 4.5D: Gene names for putative UNR-interacting proteins from SaOS-2 cells stressed with arsenite, by t-test p-value

protein	p-value	protein	p-value	protein	p-value
GN=CSDE1	0.00003	GN=WDR1	0.03440	GN=ANP32E	0.09810
GN=HSPA8	0.00070	GN=ARHGEF1	0.03500	GN=IKBKAP	0.09930
GN=TUBA1C	0.00090	GN=PHGDH	0.04030	GN=PPIA	0.10150
GN=IDH2	0.00160	GN=DHX9	0.04030	GN=P4HA2	0.10480
GN=LDB1	0.00620	GN=MCM7	0.04060	GN=TRAP1	0.10530
GN=MMS19	0.00640	GN=RPL23	0.04170	GN=EMD	0.10590
GN=RPL10	0.00660	GN=P4HB	0.04210	GN=SHMT2	0.10680
GN=STRAP	0.00750	GN=SQSTM1	0.04260	GN=ILF2	0.10980
GN=FBL	0.00810	GN=SPG20	0.04340	GN=FNDC3B	0.11070
GN=PDXDC1	0.00860	GN=TRIM25	0.04410	GN=SRSF6	0.11120
GN=RAP1B	0.00920	GN=GAPVD1	0.04420	GN=HNRNPL	0.11820
GN=RNH1	0.01050	GN=ZNFX1	0.04440	GN=ARF4	0.12170
GN=EEF2	0.01110	GN=CAD	0.04620	GN=GCN1L1	0.13190
GN=BAG2	0.01200	GN=HSPA5	0.04810	GN=RPS2	0.13360
GN=IPO4	0.01210	GN=TNRC6C	0.04910	GN=GCLM	0.13630
GN=TUBB	0.01320	GN=TUFM	0.04930	GN=HIST1H2BD	0.13750
GN=DSTN	0.01400	GN=ATAD3B	0.04980	GN=FHL2	0.13880
GN=PDS5A	0.01470	GN=ENO1	0.05050	GN=HSPBP1	0.13890
GN=EIF4G2	0.01470	GN=PGAM5	0.05070	GN=RPS3	0.14420
GN=PFKL	0.01470	GN=SLC25A3	0.05140	GN=UNC45A	0.14510
GN=RPL27A	0.01490	GN=ANKHD1	0.05470	GN=TRIM28	0.14570
GN=CFL1	0.01520	GN=SRSF3	0.05540	GN=PTPN1	0.16180
GN=PFN2	0.01600	GN=AARS	0.05620	GN=RPS6	0.16290
GN=LMO4	0.01620	GN=RPS14	0.05710	GN=UCHL1	0.16720
GN=RAB34 (NARR)	0.01840	GN=TNPO3	0.05720	GN=SRSF7	0.17890
GN=EIF2A	0.01870	GN=EIF5A	0.05720	GN=LMO7	0.18570
GN=GNAT1	0.01940	GN=RAN	0.05740	GN=RPL18	0.19660
GN=DDX39A	0.01950	GN=FXR2	0.05750	GN=EIF4A3	0.19940
GN=CTNND1	0.01990	GN=PLEKHA5	0.05800	GN=ZC3HAV1	0.20730
GN=KPNB1	0.02050	GN=ACTBL2	0.05940	GN=C4A	0.20970
GN=GNB2L1	0.02070	GN=HSPA9	0.06090	GN=LRPPRC	0.20980
GN=RPS18	0.02080	GN=SEC61A2	0.06310	GN=PLAA	0.22090
GN=H2AFV	0.02200	GN=FASN	0.06400	GN=PSMD2	0.22840
GN=PCBP1	0.02440	GN=RANBP2	0.06450	GN=SETMAR	0.23370
GN=EWSR1	0.02470	GN=RPS13	0.06570	GN=HNRNPH1	0.24060
GN=SRRM1	0.02470	GN=RPS27A	0.06670	GN=UQCRC2	0.24200
GN=SSBP3	0.02610	GN=ACADM	0.06960	GN=SSR4	0.24580
GN=ATP5A1	0.02690	GN=HSPD1	0.07050	GN=RANGAP1	0.24630
GN=TARDBP	0.02710	GN=PRKDC	0.07270	GN=WHSC1	0.24960

GN=YARS	0.02720	GN=HADHA	0.07400	GN=VIM	0.25160
GN=SF3B3	0.02750	GN=MARS	0.07520	GN=RPS8	0.25760
GN=HUWE1	0.02850	GN=HIST1H4A	0.07520	GN=TNPO1	0.27170
GN=EEF1A2	0.02890	GN=FBXO45	0.07950	GN=EPRS	0.27820
GN=ARHGEF2	0.03020	GN=RPL27	0.08030	GN=PCK2	0.31720
GN=XPOT	0.03030	GN=SUCLA2	0.08160	GN=PABPC1L	0.32110
GN=RPS16	0.03280	GN=HADHB	0.08670	GN=FLNC	0.34320
GN=ARMC6	0.03290	GN=KPNA6	0.08740	GN=CLTC	0.34360
GN=P4HA1	0.03290	GN=RPS27	0.08840	GN=NUP93	0.36060
GN=HNRNPK	0.03400	GN=TRIP6	0.08840		

N.B. Only gene names were given for the putative UNR-interacting proteins to save table space due to the large number of suggested hits. This means that some identifiers are repeated due to different isoforms being suggested. The t-test p-values are for a two tailed paired t-test without multiple testing correction and increase from top to bottom for each column, starting on the left. Yellow shading denotes proteins that were recorded in UNR samples only.

Table 4.5E: Top putative UNR interacting proteins in both plus and minus arsenite SaOS-2 samples

RL27_HUMAN 60S ribosomal protein L27 (GN=RPL27)
LMO4_HUMAN LIM domain transcription factor LMO4 (GN=LMO4)
Cluster of SSBP3_HUMAN Single-stranded DNA-binding protein 3 (GN=SSBP3)
LDB1_HUMAN Isoform 3 of LIM domain-binding protein 1 (GN=LDB1)
PGAM5_HUMAN Serine/threonine-protein phosphatase PGAM5, mitochondrial (GN=PGAM5)
Cluster of SQSTM_HUMAN Sequestosome-1 (GN=SQSTM1)
Cluster of ARHG2_HUMAN Rho guanine nucleotide exchange factor 2 (GN=ARHGEF2)
RS27_HUMAN 40S ribosomal protein S27 (GN=RPS27)
H4_HUMAN Histone H4 (GN=HIST1H4A)
Cluster of HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 (GN=HUWE1)
RS18_HUMAN 40S ribosomal protein S18 (GN=RPS18)
Cluster of SRSF6_HUMAN Serine/arginine-rich splicing factor 6 (GN=SRSF6)
RS14_HUMAN 40S ribosomal protein S14 (GN=RPS14)
EWS_HUMAN RNA-binding protein EWS (GN=EWSR1)
Cluster of C9J0D1_HUMAN Histone H2A (GN=H2AFV)
RL23_HUMAN 60S ribosomal protein L23 (GN=RPL23)
Cluster of A9Z1X7_HUMAN Serine/arginine repetitive matrix protein 1 (GN=SRRM1)
Cluster of RS3_HUMAN 40S ribosomal protein S3 (GN=RPS3)
Cluster of J3KQE5_HUMAN GTP-binding nuclear protein Ran (Fragment) (GN=RAN)
Cluster of F5H5D3_HUMAN Tubulin alpha-1C chain (GN=TUBA1C)
Cluster of CO4A_HUMAN Complement C4-A (GN=C4A)
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)
Cluster of RS6_HUMAN 40S ribosomal protein S6 (GN=RPS6)
Cluster of M0R210_HUMAN 40S ribosomal protein S16 (GN=RPS16)
Cluster of CSDE1_HUMAN Cold shock domain-containing protein E1 (GN=CSDE1)
Cluster of TBB5_HUMAN Tubulin beta chain (GN=TUBB)
Cluster of H2B1D_HUMAN Histone H2B type 1-D (GN=HIST1H2BD)
Cluster of HSP7C_HUMAN Heat shock cognate 71 kDa protein (GN=HSPA8)
SETMR_HUMAN Histone-lysine N-methyltransferase SETMAR (GN=SETMAR)
GRP78_HUMAN 78 kDa glucose-regulated protein (GN=HSPA5)
MCM7_HUMAN DNA replication licensing factor MCM7 (GN=MCM7)
STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)
P4HA2_HUMAN Prolyl 4-hydroxylase subunit alpha-2 (GN=P4HA2)

4.6.2 Discussion of putative UNR interactors found using SaOS-2 cell lysates

As with the U2OS samples, UNRIP, LDB1, NARR, HUWE1 and SSBP3 were suggested hits in both the plus and minus arsenite SaOS-2 samples (Tables 4.4E, 4.5E). The first four of these were also hits for both the plus and minus arsenite HeLa samples (Table 4.3), strongly suggesting that they may be true UNR-interacting proteins. As stated previously, SSBP3 was flagged as a putative UNR-interacting protein in both trial runs with unstressed HeLa cells (Table 4.1) and both plus and minus arsenite U2OS samples (Table 4.4E).

Two general observations were made concerning the SaOS-2 data. The first was that that the IgG samples did not contain as many proteins relative to the UNR samples, in comparison to the corresponding IgG/UNR sample pairs in the other cell types. One way this can be seen is by the number of 'yellow' putative hits for SaOS-2 compared with the other cell types. The other observation was that there were more keratins in the top ten hits by total ion current (Table 4.6).

Table 4.6: Top hits by Total Ion Intensity by cell type

HeLa	U2OS	SaOS-2
MYH9	*S. aureus* Protein A	*S. aureus* Protein A
S. aureus Protein A	MYH9	TRYP_PIG
TRYP_PIG	TRYP_PIG	KRT10
ACTBL2	ACTG1	KRT5
TPM1	KRT1	KRT1
KRT10	TPM3	K1C9
KRT2	ALBU_BOVIN	SRRM2
MYO1C	KRT14	HUWE1
MYO1E	MYL9	CSDE1
ALBU_BOVIN	K1C9	PRKDC

It was noted that UNR was one of the top ten hits for SaOS-2 but not the other cell types. Whilst it is assumed that the observations made were a perfectly representative sample of what would be seen if the experiment were to be repeated many more times, a few points must first be considered here. One point concerns myosin which is a top hit for both HeLa and U2OS, but not for SaOS-2. Exploring the overall SaOS-2 data, there was only one myosin observed as a potential interactor (MYO1B). This was very close to the bottom of the SaOS-2 list (330/398), only being observed at a low level (86361 normalised ion counts) and only in one of the twelve samples (UNR, plus arsenite, repeat 3). Interestingly, osteosarcoma cells overexpress MYH9 and it is associated with invasion and metastasis (Zhou *et al.*, 2016). Whilst this is potentially interesting, it must be stated that the final bead washing steps for the U2OS and HeLa RIP samples were kindly carried out by Dr Swagat Ray (present address University of Sheffield), whereas the author carried out the wash step for the SaOS-2 samples. The idea that the SaOS-2 washes may have been more stringent has been considered at several points over the past year. It could potentially explain the observation that the SaOS-2 IgG samples were cleaner than those of the other cell types. A highly expressed protein could be expected to bind non-specifically to IgG and even more so to a test antibody that pulls down some specific protein complexes that increase the overall protein surface available to bind to non-specific interactors. In the absence of further investigation, it must remain mere conjecture, but it is possible that myosin is a false positive.

4.7 Summary of chapter so far

It has been shown that RIP-mass spectrometry can be used to detect putative UNR-interacting proteins in HeLa cells. This was shown to be reproducible across different experiments, even where different sample preparation techniques were used. Furthermore, the detection of a group of similar proteins following arsenite treatment lends further support to the idea that the data is reproducible. Whereas it is expected from previous work in the lab that UNR localisation changes following arsenite stress, there is no reason to assume that UNR jettisons its entire set of interactors and migrates to stress granules with an entirely different set. Indeed, the working hypothesis in the lab is that at least one intracellular pool of UNR migrates to stress granules, or has stress granules form around it, whilst still being bound to the same collection of mRNA species.

It has further been shown that a proportion of the top putative UNR-interacting proteins from other cell lines are also among the top hits for HeLa cells. The fact that there are differences as well as similarities supports the idea that UNR binds a common group of proteins but also binds to others in a cell type-dependent manner.

As a result of these reassuring observations, it was decided to proceed to using more quantitative software, Progenesis (Nonlinear Dynamics, Newcastle upon Tyne, England), on the main experimental data to find final lists of putative interacting proteins that could then be exported for GO term analysis.

4.8 Progenesis

It was considered useful to carry out GO term enrichment analysis on putative UNR interacting proteins in order to see if UNR binds to functionally related groups of proteins. In order to obtain more quantitative estimates for the amount of proteins detected by mass spectroscopy, it was decided to use the Progenesis software.

Briefly, files containing raw data from the mass spectrometer for each sample to be compared were loaded into Progenesis. Default setting were employed in which Progenesis selected a sample to use as an alignment reference followed by carrying out automatic alignment of the samples. The degree of similarity between samples was provided by Progenesis, allowing for the identification and removal of potential outliers. As it was, staff in the proteomics department assured that all samples were within an acceptable range and no samples were removed. Following alignment and the quantification of potential peptides, the data were then exported to the Mascot server (Matrix Science, London, England) for peptide identification. The Mascot output files were then imported back into Progenesis and were automatically associated with their respective proteins. Low Mascot scoring proteins (score <30) and all keratins were then removed from the data. As a result of expected contamination due to the RIP protocol, pig trypsin and bovine proteins were removed, as were IgG immunoglobulins.

It was decided to use Progenesis' 'normalise to all proteins' function, as outlined at: http://nonlinear.com/progenesis/qi-for-proteomics/v2.0/faq/how-normalisation-works.aspx.

4.9 Results using Progenesis – HeLa samples

4.9.1 Progenesis analysis

In order to be confident in the identity of the peptides assigned to proteins, a series of tags were designed to label those with questionable credentials. The tags included:

1) features with no MS/MS data (i.e. things detected by the mass spectrometer, that may be peptides or contaminants, that Mascot did not recognise as a peptide under the search parameters selected – these could be modified peptides),
2) putative peptides that did not have protein IDs assigned to them,
3) putative peptides in which the minus arsenite IgG samples were highest,
4) putative peptides in which the plus arsenite IgG samples were highest,
5) putative peptides with statistical power less than 0.7.

These hits were hidden from subsequent consideration although some were later listed in the protein report where a protein was considered significant independent of them.

Suggested proteins were then considered. Those that were most abundant in one of the two IgGs groups were then tagged and removed. Finally, protein and peptide data were exported to Excel and a Progenesis protein report was generated.

4.9.2 Progenesis results for HeLa RIP dataset

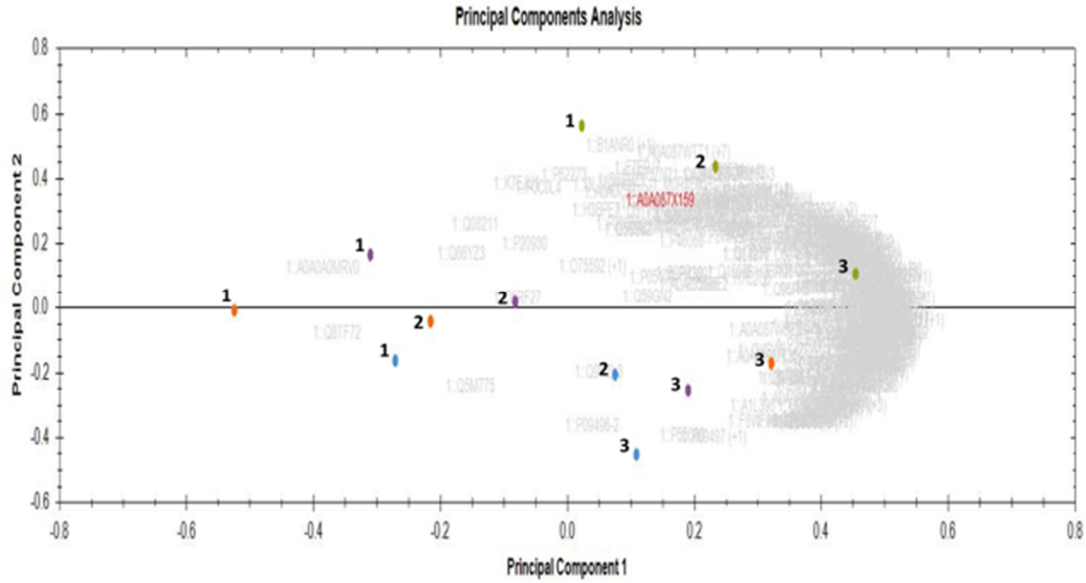
Principal component analysis can be used to locate outliers within a group of samples but it can also be used to find groupings within samples (Wold *et al.*, 1987). As it shows how the samples cluster by principal component, it is possible to confirm that known biological variability (e.g. cell type, drug treatment or whether the RIP was carried out with a test or control antibody, etc.) is responsible for the observed variation in the data.

The HeLa samples were initially imported with 361527 MS/MS spectra, of which 54635 were assigned to peptide ions. Prior to exporting to Excel, Principal Component Analysis (PCA) was carried out on the samples. This can show how samples cluster over different components of variability in the dataset and can be used to find outliers. It does that by showing whether or not samples group by sources of known biological variability (e.g. arsenite treatment or immunoprecipitating antibody) over the top principal components that, by definition, account for the largest portions of variability in the data (Wold *et al.*, 1987). In terms of the Progenesis-computed PCA analysis, there are two plots – a loadings plot that shows how the variability accounted for by each principal component is shared between each sample and a score plot that shows how the variability accounted for by each principal component is shared between each observation (e.g. peptide, protein or gene). These plots can be superimposed upon each other, as in the default images produced by Progenesis (Wold *et al.*, 1987). For further consideration of the use of PCA as a tool (focusing on loadings plots), see section 4.2 below.

It was decided to remove proteins with an ANOVA p-value over 0.05. This was for several reasons. Firstly, it improved the clustering of the repeats over the first two principle components of the PCA which also accounted for more of the variance (Figure 4.13). Secondly, it was decided

more important to detect a small number of UNR-interacting proteins with greater confidence than to detect a larger number that was more likely to contain false positives. ANOVA detects differences between group means, thereby ignoring the pairing in the data (the UNR and IgG samples from a biological repeat were pulled down from exactly the same lysate and the plus and minus arsenite treatment lysates from a particular repeat only differed in treatment one hour prior to harvesting). This makes it more at risk of outliers skewing means and generating false positives than paired t-tests. Paired t-tests look for differences in the differences between the UNR and IgG samples for each repeat rather than the difference in their means. This is more powerful and can better detect hits with small but consistent increases in UNR over IgG. By using both tests, it makes it more likely that any hits that are detected will have both significantly higher UNR means than IgG means and also show a general trend of UNR being higher than IgG.

A



B

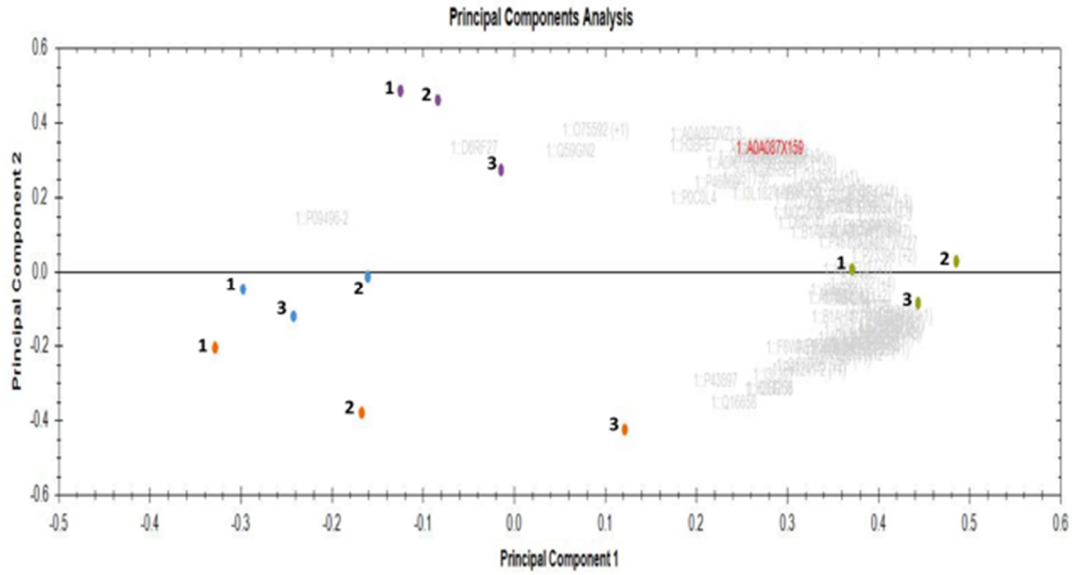


Figure 4.13: Progenesis-generated PCA plots for HeLa data. Proteins with all ANOVA p-values (A) or only those with ANOVA p-values under 0.05 (B) were included. See text for further information how the data were processed. Numbers are associated with the days upon which the lysates were made (i.e. the repeats). Blue = no arsenite, IgG; purple = no arsenite, UNR; orange = plus arsenite, IgG; green = plus arsenite, UNR. In terms of variance accounted for – A: (PC1 = 59.09%, PC2 = 15.04%); B: (PC1 = 57.61%, PC2 = 23.05%). The grey writing in the background makes up the protein score plot (see text for discussion of score plots). The red writing refers to an automatically highlighted protein ID, selected alphabetically.

Having exported the data into Excel, paired t-tests were carried out on the plus and minus arsenite samples separately. This suggested 25 putative UNR interacting proteins from the minus arsenite samples when the cut off p-value was set at 0.1 (Table 4.7B), the top ten putative UNR interactors being recorded by p-value in Table 4.7A. Meanwhile, 67 putative hits were suggested for the plus arsenite data (Table 4.8B). The top ten of these, by t test p value, are presented in Table 4.8A.

Table 4.7A: Top 10 putative HeLa minus arsenite UNR interacting proteins by t-test p-value (using Progenesis data)

Protein	HeLa1	HeLa2	HeLa3	HeLa4	HeLa5	HeLa6	p-value
HNRNPR	1484	4209	1090	18582	20973	17261	0.0003
UNR (E9PLT0)	675629	1432042	978551	8203345	8312966	7468645	0.0019
SSBP2	9576	9430	9565	2855595	2442834	2576302	0.0022
UNR (CSDE1)	617677	1294118	892943	5277410	5125360	4775903	0.0042
SSBP3	13708	20111	42631	3465982	3242002	4053717	0.0043
P2RY12	2339	3920	1969	21546	27268	26049	0.0046
C4A	5006	1456	0	164581	127163	148054	0.0047
NARR	14859	16754	23707	1760160	2189489	2435130	0.0084
LDB1	26760	85648	60981	2145634	1936377	2763112	0.0125
RPS18	2392444	3641255	1930012	3150191	4154056	2662038	0.0133

N.B. IgG samples shaded lilac, repeat 1-3 running left to right (HeLa1-HeLa3). The corresponding UNR samples are shaded green and also run left to right (HeLa4-HeLa6). The t-test p-values are for a two tailed paired t-test without multiple testing correction. Two UNR proteins were highlighted in the top ten; the respective Progenesis-derived identifiers are provided in parentheses.

Table 4.7B: Putative UNR interacting proteins by t-test from HeLa minus arsenite samples (using Progenesis data with an unadjusted p-value cut-off of p=0.05)

Protein	p-value
HNRPR_HUMAN Heterogeneous nuclear ribonucleoprotein R (GN=HNRNPR)	0.00026
E9PLT0_HUMAN Cold shock domain-containing protein E1 (GN=CSDE1)	0.00188
A0A087X159_HUMAN Single-stranded DNA-binding protein 2 (GN=SSBP2)	0.00215
CSDE1_HUMAN Cold shock domain-containing protein E1 (GN=CSDE1)	0.00419
A0A087WVT6_HUMAN Single-stranded DNA-binding protein 3 (GN=SSBP3)	0.00430
P2Y12_HUMAN P2Y purinoceptor 12 (GN=P2RY12)	0.00463
CO4A_HUMAN Complement C4-A (GN=C4A)	0.00470
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)	0.00842
LDB1_HUMAN LIM domain-binding protein 1 (GN=LDB1)	0.01253
RS18_HUMAN 40S ribosomal protein S18 (GN=RPS18)	0.01329
STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)	0.01465
TRI25_HUMAN E3 ubiquitin/ISG15 ligase TRIM25 (GN=TRIM25)	0.02343
Q05BK6_HUMAN Protein TFG (GN=TFG)	0.02749
LMO4_HUMAN LIM domain transcription factor LMO4 (GN=LMO4)	0.03034
A0A087WZL3_HUMAN Tyrosine-protein kinase receptor (GN=ALK)	0.03048
CSDE1_HUMAN Isoform 3 of Cold shock domain-containing protein E1 (GN=CSDE1	0.04178
E9PMM9_HUMAN 40S ribosomal protein S2 (Fragment) (GN=RPS2)	0.05057
FSCN1_HUMAN Fascin (GN=FSCN1)	0.05879
SYAC_HUMAN Alanine--tRNA ligase, cytoplasmic (GN=AARS)	0.05934
SRSF3_HUMAN Serine/arginine-rich splicing factor 3 (GN=SRSF3)	0.06548
I3L182_HUMAN Serine/arginine repetitive matrix protein 2 (Fragment) (GN=SRRM2)	0.07263
E9PQD7_HUMAN 40S ribosomal protein S2 (GN=RPS2)	0.07390
E7EMC7_HUMAN Sequestosome-1 (GN=SQSTM1)	0.07701
A0A087X1I3_HUMAN Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (GN=SDHA)	0.08195
RINI_HUMAN Ribonuclease inhibitor (GN=RNH1)	0.09982

Table 4.8A: Top ten putative HeLa plus arsenite UNR interacting proteins by t-test p-value (using Progenesis data)

Description	HeLa7	HeLa8	HeLa9	HeLa10	HeLa11	HeLa12	p-value
RPS23	189587	307227	608349	683612	812599	1136187	0.0004
SSBP3	21971	20181	133066	1972922	2219222	2205696	0.0012
ALK	262326	240564	249102	398416	359831	361803	0.0032
NARR	7306	8763	34080	1671992	1540368	1912650	0.0036
EIF2A	1182	4059	20313	68570	82766	103307	0.0037
RPL38	60934	100721	143662	240446	332399	363123	0.0056
SSBP2	13362	5506	8641	2549008	2247058	1897033	0.0070
RNH1 (E9PIK5)	68991	151294	1121271	944800	1267502	1973206	0.0078
LMO4	8132	13706	3252	1083616	1156444	1436151	0.0080
RNH1 (RNH1)	83883	201182	1698218	1370368	1955670	3032133	0.0102

N.B. IgG samples shaded lilac, repeat 1-3 running left to right (HeLa7-HeLa9). The corresponding UNR samples are shaded green and also run left to right (HeLa10-HeLa12). The t-test p-values are for a two tailed paired t-test without multiple testing correction. Two RNH1 proteins were highlighted in the top ten; the respective Progenesis-derived identifiers are provided in parentheses.

Table 4.8B: Putative UNR interacting proteins by t-test from HeLa plus arsenite samples (using Progenesis data with an unadjusted p-value cut-off of p=0.05)

Protein	p-value
D6RD47_HUMAN 40S ribosomal protein S23 (GN=RPS23)	0.00038
A0A087WVT6_HUMAN Single-stranded DNA-binding protein 3 (GN=SSBP3)	0.00119
A0A087WZL3_HUMAN Tyrosine-protein kinase receptor (GN=ALK)	0.00321
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)	0.00355
F8WAE5_HUMAN Eukaryotic translation initiation factor 2A (GN=EIF2A)	0.00369
J3KSP2_HUMAN 60S ribosomal protein L38 (Fragment) (GN=RPL38)	0.00557
A0A087X159_HUMAN Single-stranded DNA-binding protein 2 (GN=SSBP2)	0.00702
E9PIK5_HUMAN Ribonuclease inhibitor (Fragment) (GN=RNH1)	0.00783
LMO4_HUMAN LIM domain transcription factor LMO4 (GN=LMO4)	0.00802
RINI_HUMAN Ribonuclease inhibitor (GN=RNH1)	0.01024
E9PJD9_HUMAN 60S ribosomal protein L27a (GN=RPL27A)	0.01175
PCBP1_HUMAN Poly(rC)-binding protein 1 (GN=PCBP1)	0.01204
F5H2T0_HUMAN Elongator complex protein 1 (GN=IKBKAP)	0.01228
LDB1_HUMAN LIM domain-binding protein 1 (GN=LDB1)	0.01307
CO4A_HUMAN Complement C4-A (GN=C4A)	0.01317
FAS_HUMAN Fatty acid synthase (GN=FASN)	0.01320
I3L182_HUMAN Serine/arginine repetitive matrix protein 2 (Fragment) (GN=SRRM2)	0.01602
J3KT73_HUMAN 60S ribosomal protein L38 (GN=RPL38)	0.01861
H3BRU6_HUMAN Poly(rC)-binding protein 2 (Fragment) (GN=PCBP2)	0.02154
F6RFD5_HUMAN Destrin (GN=DSTN)	0.02474
D6R9Z1_HUMAN Guanine nucleotide-binding protein subunit beta-2-like 1 (Fragment) (GN=GNB2L1)	0.02519
E9PMM9_HUMAN 40S ribosomal protein S2 (Fragment) (GN=RPS2)	0.02536
SYAC_HUMAN Alanine--tRNA ligase, cytoplasmic (GN=AARS)	0.02577
E9PQD7_HUMAN 40S ribosomal protein S2 (GN=RPS2)	0.02614
HNRPR_HUMAN Heterogeneous nuclear ribonucleoprotein R (GN=HNRNPR)	0.02710
A0A087WYT3_HUMAN Prostaglandin E synthase 3 (GN=PTGES3)	0.02935
P2Y12_HUMAN P2Y purinoceptor 12 (GN=P2RY12)	0.02989
CSDE1_HUMAN Isoform 3 of Cold shock domain-containing protein E1 (GN=CSDE1)	0.03023
A0A075B730_HUMAN Epiplakin (GN=EPPK1)	0.03178
E9PPT0_HUMAN 40S ribosomal protein S2 (GN=RPS2)	0.03203
Q5T6W2_HUMAN Heterogeneous nuclear ribonucleoprotein K (Fragment) (GN=HNRNPK)	0.03302
B4DXZ6_HUMAN Fragile X mental retardation syndrome-related protein 1 (GN=FXR1)	0.03337
H0YEU2_HUMAN 40S ribosomal protein S3 (Fragment) (GN=RPS3)	0.03337
A0A0D9SFL3_HUMAN RNA-binding protein EWS (GN=EWSR1)	0.03351

EF2_HUMAN Elongation factor 2 (GN=EEF2)	0.03383
TADBP_HUMAN TAR DNA-binding protein 43 (GN=TARDBP)	0.03518
CSDE1_HUMAN Cold shock domain-containing protein E1 (GN=CSDE1)	0.03577
SND1_HUMAN Staphylococcal nuclease domain-containing protein 1 (GN=SND1)	0.03618
B1AH49_HUMAN Sulfurtransferase (GN=MPST)	0.03642
MYCB2_HUMAN E3 ubiquitin-protein ligase MYCBP2 (GN=MYCBP2)	0.03940
ZCCHV_HUMAN Zinc finger CCCH-type antiviral protein 1 (GN=ZC3HAV1)	0.04165
E9PFP8_HUMAN Poly(rC)-binding protein 3 (GN=PCBP3)	0.04277
A0A087X1I3_HUMAN Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (GN=SDHA)	0.04313
E9PLT0_HUMAN Cold shock domain-containing protein E1 (GN=CSDE1)	0.04357
STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)	0.04656
EPIPL_HUMAN Epiplakin (GN=EPPK1)	0.04902
G3V1A4_HUMAN Cofilin 1 (Non-muscle), isoform CRA_a (GN=CFL1)	0.04947
I3L397_HUMAN Eukaryotic translation initiation factor 5A (Fragment) (GN=EIF5A)	0.05004
IF2G_HUMAN Eukaryotic translation initiation factor 2 subunit 3 (GN=EIF2S3)	0.05009
NSUN2_HUMAN tRNA (cytosine(34)-C(5))-methyltransferase (GN=NSUN2)	0.05026
HNRPK_HUMAN Heterogeneous nuclear ribonucleoprotein K (GN=HNRNPK)	0.05028
RS18_HUMAN 40S ribosomal protein S18 (GN=RPS18)	0.05052
K7EJT5_HUMAN 60S ribosomal protein L22 (Fragment) (GN=RPL22)	0.05817
XP32_HUMAN Skin-specific protein 32 (GN=XP32)	0.05887
M0QXS5_HUMAN Heterogeneous nuclear ribonucleoprotein L (Fragment) (GN=HNRNPL)	0.06395
TRI25_HUMAN E3 ubiquitin/ISG15 ligase TRIM25 (GN=TRIM25)	0.06435
TRIP6_HUMAN Thyroid receptor-interacting protein 6 (GN=TRIP6)	0.06435
RS3_HUMAN 40S ribosomal protein S3 (GN=RPS3)	0.06822
I3L3P7_HUMAN 40S ribosomal protein S15a (GN=RPS15A)	0.07117
RS10_HUMAN 40S ribosomal protein S10 (GN=RPS10)	0.07213
M0QZN2_HUMAN 40S ribosomal protein S5 (GN=RPS5)	0.07301
A0A087WZ27_HUMAN Zinc finger protein 90 (GN=ZNF90)	0.07448
X6RLN4_HUMAN La-related protein 4 (Fragment) (GN=LARP4)	0.07573
HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 (GN=HUWE1)	0.08909
MCTP2_HUMAN Multiple C2 and transmembrane domain-containing protein 2 (GN=MCTP2)	0.08947
H0YL92_HUMAN Importin-4 (GN=IPO4)	0.09628
FSCN1_HUMAN Fascin (GN=FSCN1)	0.09700

The peptides of some of the suggested hits were analysed to check that they were all following the same general trend in terms of being higher in the group that was suggested to be higher (e.g. UNR pulldowns from the HeLa plus arsenite lysates) and similar in every repeat (Figure 4.14). Such a pattern would be in keeping with a genuine hit, whereas a more random pattern would suggest that some of the peptides had been incorrectly assigned and that the protein was, therefore, a more questionable hit.

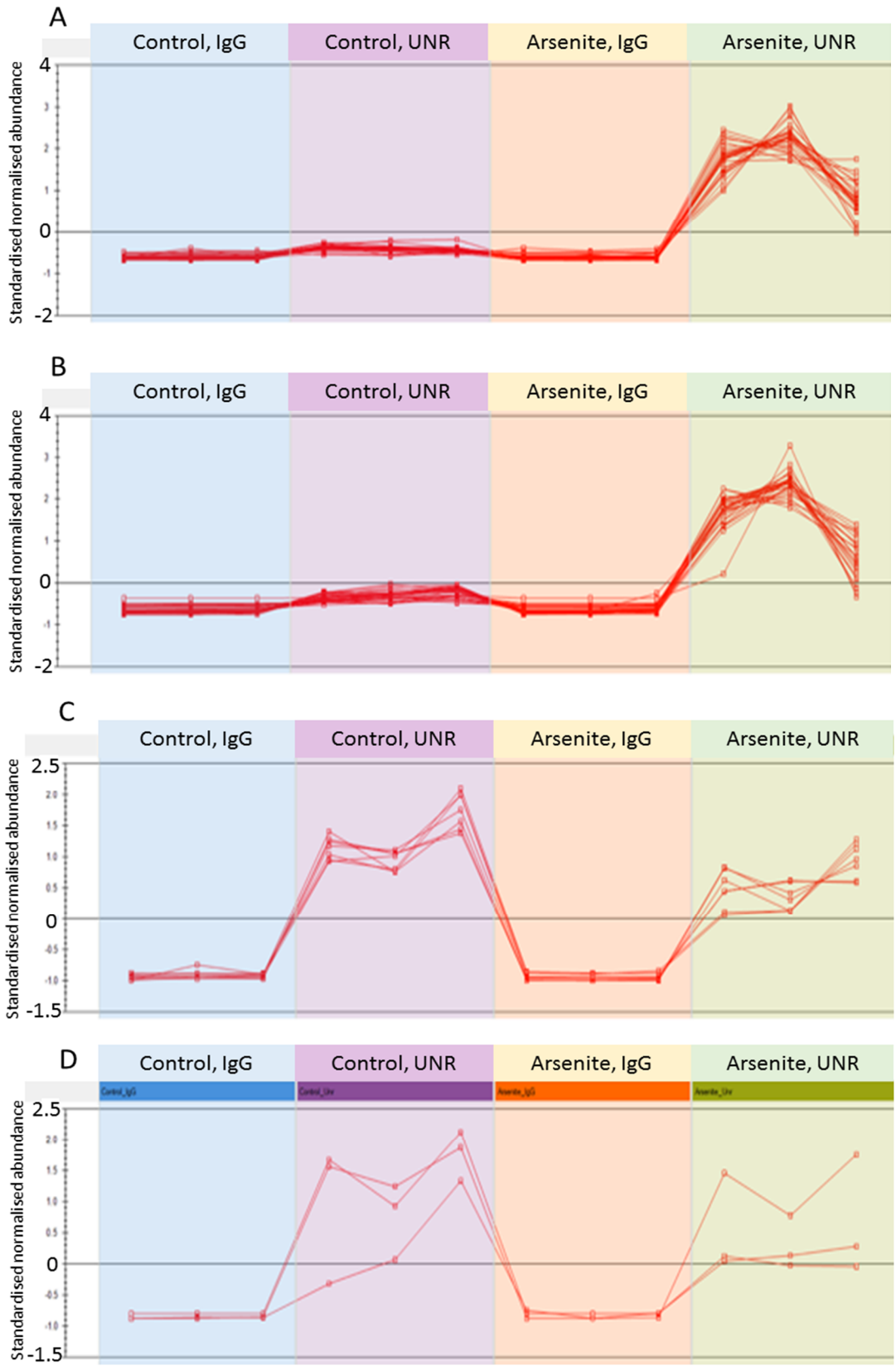


Figure 4.14: Standardised normalised abundances for peptides assigned to specific proteins (A=UNR, B=UNRIP, C = LDB1, D=SSBP3) from the HeLa samples. The repeats in the blue column are IgG (minus arsenite), the repeats in the purple column are UNR (minus arsenite), the repeats in the orange column are IgG (plus arsenite) and the repeats in the green column are UNR (plus arsenite).

The peptides for UNR and LDB1 followed approximately the same pattern as each other, supporting the idea that they may come from the same protein. One peptide for UNRIP (most pronounced in green, UNR plus arsenite, panel) and one for SSBP3 (low in purple, UNR minus arsenite, and green, UNR plus arsenite, panels), however, seemed to follow a different pattern to the others (Figure 4.14). This implies that those peptides may have been incorrectly assigned to the protein.

Finally, fold change data was considered. A fold change minimum of 10 was set as a significance cut off point (again only considering the data with ANOVA p-value of 0.05 or less). This gave a set of suggested UNR-interacting proteins in unstressed HeLa (Table 4.9B) and in arsenite stressed HeLa (Table 4.10B). The top ten putative UNR-interacting proteins in each case are presented in Tables 4.9A and 4.10A, respectively.

Table 4.9A: Top 10 putative HeLa minus arsenite UNR interacting proteins by mean UNR/IgG ratios (using Progenesis data)

Description	HeLa1	HeLa2	HeLa3	HeLa4	HeLa5	HeLa6	Ratio
ILF2	0	0	0	10400	11648	1425	∞
LMO4	5126	5328	2150	1875371	1747171	2977132	524
SSBP2	9576	9430	9565	2855595	2442834	2576302	276
TAF15	1324	264	1686	735092	131738	11583	268
FUS	7301	420	919	1244136	272313	28064	179
SSBP3	13708	20111	42631	3465982	3242002	4053717	141
NARR	14859	16754	23707	1760160	2189489	2435130	115
C4A	5006	1456	0	164581	127163	148054	68
HUWE1	859	514	11976	75310	118533	430660	47
LDB1	26760	85648	60981	2145634	1936377	2763112	39

N.B. IgG samples shaded lilac, repeat 1-3 running left to right (HeLa1-HeLa3). The corresponding UNR samples are shaded green and also run left to right (HeLa4-HeLa6). Ratios are mean(UNR)/mean(IgG) and all infinities were ignored unless every UNR value was greater than zero. The yellow shading highlights a protein absent in every IgG sample but present in every UNR sample.

Table 4.9B: Putative HeLa minus arsenite UNR interacting proteins by mean UNR/IgG ratios (using Progenesis data)

Description	Ratio
B4DY09_HUMAN Interleukin enhancer-binding factor 2 (GN=ILF2)	∞
LMO4_HUMAN LIM domain transcription factor LMO4 (GN=LMO4)	524
A0A087X159_HUMAN Single-stranded DNA-binding protein 2 (GN=SSBP2)	276
RBP56_HUMAN TATA-binding protein-associated factor 2N (GN=TAF15)	268
H3BPE7_HUMAN RNA-binding protein FUS (GN=FUS)	179
A0A087WVT6_HUMAN Single-stranded DNA-binding protein 3 (GN=SSBP3)	141
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)	115
CO4A_HUMAN Complement C4-A (GN=C4A)	68
HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 (GN=HUWE1)	47
LDB1_HUMAN LIM domain-binding protein 1 (GN=LDB1)	39
Q05BK6_HUMAN Protein TFG (GN=TFG)	32
MYCB2_HUMAN E3 ubiquitin-protein ligase MYCBP2 (GN=MYCBP2)	21
I3L182_HUMAN Serine/arginine repetitive matrix protein 2 (Fragment) (GN=SRRM2)	18
STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)	16
D6R9Z1_HUMAN Guanine nucleotide-binding protein subunit beta-2-like 1 (Fragment) (GN=GNB2L1)	14
SQSTM_HUMAN Sequestosome-1 (GN=SQSTM1)	13
A0A075B7D9_HUMAN TATA-binding protein-associated factor 2N (GN=TAF15)	12

N.B. The yellow shading highlights a protein absent in every IgG sample but present in every UNR sample – this therefore produces an infinite UNR/IgG ratio.

Table 4.10A: Top 10 putative HeLa plus arsenite UNR interacting proteins by ratio (using Progenesis data)

Description	HeLa7	HeLa8	HeLa9	HeLa10	HeLa11	HeLa12	Ratio
RANGAP1	126	0	0	11726	50295	4087	526
TAF15	64	0	2142	594692	217939	20502	378
FUS	2940	0	257	690464	339086	43128	335
SSBP2	13362	5506	8641	2549008	2247058	1897033	243
LMO4	8132	13706	3252	1083616	1156444	1436151	147
TFG	781	1316	11614	309614	925356	419564	121
NARR	7306	8763	34080	1671992	1540368	1912650	102
UNR	597845	805286	1371284	75654890	90439807	41375065	75
UNRIP	540420	543636	950744	55449907	65199826	28849813	73
P2RY12	368	1453	5270	146665	163822	91204	57

N.B. IgG samples shaded lilac, repeat 1-3 running left to right (HeLa7-HeLa9). The corresponding UNR samples are shaded green and also run left to right (HeLa10-HeLa12). Ratios are mean(UNR)/mean(IgG) and all infinities were ignored unless every UNR value was greater than zero.

Table 4.10B: Putative HeLa plus arsenite UNR interacting proteins by mean UNR/IgG ratios (using Progenesis data)

Description	Ratio
RAGP1_HUMAN Ran GTPase-activating protein 1 (GN=RANGAP1)	526
RBP56_HUMAN TATA-binding protein-associated factor 2N (GN=TAF15)	378
H3BPE7_HUMAN RNA-binding protein FUS (GN=FUS)	335
A0A087X159_HUMAN Single-stranded DNA-binding protein 2 (GN=SSBP2)	243
LMO4_HUMAN LIM domain transcription factor LMO4 (GN=LMO4)	147
Q05BK6_HUMAN Protein TFG (GN=TFG)	121
NARR_HUMAN Ras-related protein RAB34, isoform NARR (GN=RAB34)	102
E9PLT0_HUMAN Cold shock domain-containing protein E1 (GN=CSDE1)	75
STRAP_HUMAN Serine-threonine kinase receptor-associated protein (GN=STRAP)	73
P2Y12_HUMAN P2Y purinoceptor 12 (GN=P2RY12)	57
CSDE1_HUMAN Cold shock domain-containing protein E1 (GN=CSDE1)	56
LDB1_HUMAN LIM domain-binding protein 1 (GN=LDB1)	52
CSDE1_HUMAN Isoform 3 of Cold shock domain-containing protein E1 (GN=CSDE1)	52
SQSTM_HUMAN Sequestosome-1 (GN=SQSTM1)	41
B4DY09_HUMAN Interleukin enhancer-binding factor 2 (GN=ILF2)	39
A0A087WVT6_HUMAN Single-stranded DNA-binding protein 3 (GN=SSBP3)	37
B1ANR0_HUMAN Polyadenylate-binding protein (GN=PABPC4)	31
E7EMC7_HUMAN Sequestosome-1 (GN=SQSTM1)	28
A0A087WTT1_HUMAN Polyadenylate-binding protein (GN=PABPC1)	25
A0A075B7D9_HUMAN TATA-binding protein-associated factor 2N (GN=TAF15)	18
RBP2_HUMAN E3 SUMO-protein ligase RanBP2 (GN=RANBP2)	16
HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 (GN=HUWE1)	14
I3L182_HUMAN Serine/arginine repetitive matrix protein 2 (Fragment) (GN=SRRM2)	12
CO4A_HUMAN Complement C4-A (GN=C4A)	12
MCTP2_HUMAN Multiple C2 and transmembrane domain-containing protein 2 (GN=MCTP2)	11

4.9.3 Discussion of putative UNR-interacting proteins in HeLa (Progenesis)

By eye, the top ten HeLa minus arsenite samples by p-value were encouraging (Table 4.7). The differences between the IgG and UNR samples were quite pronounced. There were two entries associated with UNR. As stated previously, it is unlikely that E9PLTO is physiologically relevant and was likely to have been flagged up as a result of parsimony-based algorithms assigning relatively more peptides to a shorter entry than for a larger one. It was reassuring to see that it is not the only entry in the top ten. Furthermore, three of the top ten hits have been mentioned previously (LDB1, NARR and SSBP3). The last of these was present in the company of the functionally related SSBP2. The top hit, showing more than a tenfold increase in UNR samples over IgG samples is HNRNPR. This, and the ribosomal protein S18, were also in keeping with the assumption that UNR interacts with and modulates the translation of mRNAs. The other two hits, a P2Y receptor and a complement protein are interesting but their relevance were not immediately apparent.

There was some overlap between the top ten putative UNR-interacting proteins in non-arsenite treated HeLa cells by t-test p-value (Table 4.7) and by UNR/IgG ratio (Table 4.9). SSBP2 and SSBP3 were both present, as was NARR, LDB1 and the complement protein C4-A. LMO4 had the highest non-infinite ratio, giving strong support to the idea that this protein is a true UNR-interactor. FUS and the related TAF15 are members of the TET family that has a variety of cellular functions (Takahama *et al.*, 2008). These include the control of alternative splicing and transcription (Ishigaki *et al.*, 2012), the regulation of a variety of different protein levels in neurons (Ibrahim *et al.*, 2013) and they are involved in various pathologies (discussed in Schwartz et al. 2015). Interestingly, although the HeLa cells in this case were unstressed, both FUS and TAF15 are known to migrate to stress granules in stressed cells (Blechingberg *et al.*, 2012). Another interesting protein in the top ten UNR/IgG ratio list was the ubiquitin E3 ligase HUWE1. HUWE1, which has been linked to multiple cancers, has multiple aliases (succinctly reviewed in Choe et al. 2016). It has been reported to promote the restart of replication at stalled replication forks (Choe *et al.*, 2016). It is also linked to X-linked intellectual impairment (Orivoli *et al.*, 2016). Intriguingly, the Huwe1 gene is host to miR-98 and the expression of the pair was shown to be positively correlated (Xu *et al.*, 2015). Furthermore, miR-98 was shown to downregulate the effector caspase, Caspase 3 (Xu *et al.*, 2015). Fas expression was also shown to be downregulated by miR-98 (Wang *et al.*, 2011). Whilst not directly related to the putative interaction between UNR and HUWE1, *per* se, these findings regarding miR-98 imply that

HUWE1 gene can have an oncogenic function beyond the HUWE1 protein by reducing the ability of the cell to undergo apoptosis. The top hit by ratio was Interleukin enhancer-binding factor 2. It was not present in the IgG samples and was by far the lowest of the top ten in terms of the amount being bound to UNR. It is also known as NF45 and can reduce the production of mature miRNAs in conjunction with ILF3 (also known as NF90) (Sakamoto *et al.*, 2009). It is upregulated in non-small cell lung cancer, where it is correlated with a poor prognosis (Ni *et al.*, 2015), and is involved in splicing and in the DNA damage response in 1q21-amplified multiple myeloma (Marchesini *et al.*, 2014). It was noted that multiple myeloma related proteins had been discussed previously (Figure 4.11).

The top ten putative UNR-interacting proteins from arsenite-treated HeLa cell lysates by t-test p-value were also of interest (Table 4.8). As well as many of the common hits (e.g. SSBP3, NARR, SSBP2 and LMO4), there were five proteins directly linked to RNA. These were the ribosomal proteins S23 and L38, eukaryotic translation initiation factor 2A and two entries for Ribonuclease inhibitor. It was worth noting that the recombinant RNase OUT ribonuclease inhibitor (Invitrogen) had been added to the samples prior to the RIP step. That observation led to two assumptions; either the highlighted RNase inhibitor was the exogenous RNase Out or, alternatively, the human protein that was suggested was genuinely present. Given that the Mascot searches used a human database, either option seems reasonable. As RNase inhibitors could be associated with the RNA species that were part of RNP complexes with UNR, that could imply that UNR was associated with more RNA species, or had at least pulled down more RNA molecules, in the plus arsenite samples than in the minus arsenite samples. The other hit, anaplastic lymphoma receptor tyrosine kinase, is interesting due to it being a membrane-bound protein. It is involved in neuronal differentiation (Gouzi *et al.*, 2005) and, unusually, it is activated in a ligand-independent manner (Deuel, 2013). It can become fused to a variety of proteins by genetic translocations, these include MYH9 which had been suggested as the top hit by total ion intensity in Scaffold (Table 4.6) (Lamant *et al.*, 2003).

The top ten putative UNR-interacting proteins in arsenite-treated HeLa cells by UNR/IgG ratio contained many of the proteins already discussed (Table 4.10). These included; TAF15, FUS, SSBP2, LMO4, NARR, UNR itself and UNRIP. The remaining proteins included Ran GTPase-activating protein 1 (RANGAP1). The canonical function of a GAP protein is to activate the GTPase activity of small monomeric G proteins, thereby causing them to cleave their associated GTP and functioning as an ‘off switch’. The rangap1 protein was shown to function in the

regulation of kinetochores and knocking it down prevented activation of the spindle checkpoint (Arnaoutov & Dasso, 2003). SUMOylation with SUMO1 was involved with causing RANGAP1 to migrate to the nuclear pore (Matunis *et al.*, 1998). RANGAP1 has also been associated with a number of cancers including diffuse large B cell lymphoma (Chang *et al.*, 2013). The 'TRK-fused gene' protein is involved in regulating endoplasmic reticulum structure and protein release (Beetz *et al.*, 2013). It is interesting in that it is a potential fusion partner for one of the proteins noted above as a top hit by t-test from lysate made from arsenite-treated HeLa cells (ALK, Table 4.8A). The last remaining suggested top hit was the same P2Y receptor (P2RY12) as was suggested in the untreated HeLa samples by t-test. It is involved in platelet activation (discussed in Dorsam & Kunapuli 2004) and is the target of clopidogrel (Savi *et al.*, 2001).

4.9.4 Progenesis results for HeLa RIP dataset exported to AmiGO 2

The plus and minus arsenite hits suggested by Progenesis were exported separately to AmiGO 2 version 2.3.2 at http://amigo.geneontology.org/rte.

To do this, all proteins that were considered significant (i.e. having a t-test p-value under 0.1 or having a UNR/IgG ratio of over ten) were exported to Excel. Gene names ('GN=') were then isolated (except for NARR which is known to function differently to Rab34), duplicates were removed and the remaining gene names were entered into the tool and searched using the following annotation data sets:

- GO biological process complete,
- GO molecular function complete,
- GO cellular component complete.

The GO term searches were carried out with the parameters in Table 4.11.

Table 4.11: AmiGO 2 search parameters

Analysis Type:	PANTHER Overrepresentation Test (release 20160321)
Annotation Version and Release Date:	GO Ontology database Released 2016-04-23
Analysed List:	upload_1 (Homo sapiens)
Reference List:	Homo sapiens (all genes in database)
Bonferroni correction:	TRUE

4.9.5 GO term results for HeLa minus arsenite

A total of 34 individual proteins were entered into AmiGO 2 and all of them were recognised by the tool expect for NARR. Faced with the decision of including an entirely different protein (Rab34) or excluding NARR altogether, the latter option was considered more appropriate.

A Bonferroni multiple testing corrected p-value of 0.05 was chosen as the significance cut off point and the top ten enriched GO terms by adjusted p-value, where possible, were recorded in Table 4.12 (Table 4.12A = biological process, Table 4.12B = molecular function, Table 4.12C = cellular component).

Table 4.12A: Top ten enriched biological process GO terms, by Bonferroni adjusted p-value (HeLa minus arsenite samples)

GO biological process complete	Ref	Obs	Exp	FE	p-value
mRNA metabolic process (GO:0016071)	625	13	1	13.1	4.34E-08
RNA processing (GO:0006396)	930	13	1.5	8.82	5.74E-06
mRNA catabolic process (GO:0006402)	206	7	0.3	21.4	2.62E-04
viral transcription (GO:0019083)	119	6	0.2	31.8	2.80E-04
viral gene expression (GO:0019080)	131	6	0.2	28.9	4.91E-04
RNA catabolic process (GO:0006401)	235	7	0.4	18.8	6.37E-04
multi-organism metabolic process (GO:0044033)	159	6	0.3	23.8	1.52E-03
macromolecule catabolic process (GO:0009057)	1025	11	1.6	6.77	2.40E-03
selenocysteine metabolic process (GO:0016259)	88	5	0.1	35.8	2.40E-03
cellular macromolecule catabolic process (GO:0044265)	807	10	1.3	7.82	2.58E-03

N.B. Ref = the number of proteins with a given annotation from the entire human reference list used (20814 entries), Obs = number of proteins observed with specified annotation, Exp = number of proteins expected by chance alone based on the given sample size, FE = fold enrichment (i.e. Obs/Exp).

Table 4.12B: All enriched molecular function GO terms, by Bonferroni adjusted p-value (HeLa minus arsenite samples)

GO molecular function complete	Ref	Obs	Exp	FE	p-value
RNA binding (GO:0003723)	1718	21	2.7	7.71	6.40E-12
poly(A) RNA binding (GO:0044822)	1165	18	1.9	9.75	3.61E-11
nucleic acid binding (GO:0003676)	4514	26	7.2	3.63	1.24E-08
heterocyclic compound binding (GO:1901363)	6512	28	10	2.71	7.92E-07
organic cyclic compound binding (GO:0097159)	6581	28	10	2.68	1.04E-06
mRNA binding (GO:0003729)	160	7	0.3	27.6	1.54E-05
protein binding (GO:0005515)	10761	29	17	1.7	3.44E-02

N.B. Ref = the number of proteins with a given annotation from the entire human reference list used (20814 entries), Obs = number of proteins observed with specified annotation, Exp = number of proteins expected by chance alone based on the given sample size, FE = fold enrichment (i.e. Obs/Exp). As there were only 7 enriched GO terms that passed the Bonferroni multiple testing correction, all significantly enriched GO terms were recorded.

Table 4.12C: Top ten enriched cellular component GO terms, by Bonferroni adjusted p-value (HeLa minus arsenite samples)

GO cellular component complete	Ref	Obs	Exp	FE	p-value
cytosolic small ribosomal subunit (GO:0022627)	41	5	0.1	76.9	8.68E-06
ribonucleoprotein complex (GO:1990904)	750	11	1.2	9.25	1.59E-05
intracellular ribonucleoprotein complex (GO:0030529)	750	11	1.2	9.25	1.59E-05
small ribosomal subunit (GO:0015935)	70	5	0.1	45.1	1.22E-04
macromolecular complex (GO:0032991)	4651	21	7.4	2.85	5.60E-04
cytosolic ribosome (GO:0022626)	110	5	0.2	28.7	1.12E-03
extracellular exosome (GO:0070062)	2725	16	4.3	3.7	1.21E-03
extracellular vesicle (GO:1903561)	2738	16	4.3	3.69	1.29E-03
extracellular organelle (GO:0043230)	2740	16	4.3	3.68	1.31E-03
organelle (GO:0043226)	12818	32	20	1.57	3.14E-03

N.B. Ref = the number of proteins with a given annotation from the entire human reference list used (20814 entries), Obs = number of proteins observed with specified annotation, Exp = number of proteins expected by chance alone based on the given sample size, FE = fold enrichment (i.e. Obs/Exp).

4.9.6 Consideration of GO term results for HeLa minus arsenite

The top hits were largely related to RNA (Tables 4.12A, 4.12B, 4.12C). As UNR is known to be an RNA-binding protein, this is not unexpected. Although it is already known, it would seem reasonable to conclude from this data alone that UNR is likely to be intimately involved with RNA.

Other than RNA related GO terms, viral gene expression and selenocysteine metabolic processes were also noted from the biological process data (Table 4.12A). Extracellular exosome/vesicle/organelle was also noted (Table 4.12C).

4.9.7 GO term results for HeLa plus arsenite

Other than NARR, which was removed from the data as above, a total of 67 individual proteins were entered into AmiGO 2 and all of them were recognised by the tool. A Bonferroni multiple testing corrected p-value of 0.05 was chosen as the significance cut off point and the top ten enriched GO terms by adjusted p-value were recorded in Table 4.13 (Table 4.13A = biological process, Table 4.13B = molecular function, Table 4.13C = cellular component).

Table 4.13A: Top ten enriched biological process GO terms, by Bonferroni adjusted p-value (HeLa plus arsenite samples)

GO biological process complete	Ref	Obs	Exp	FE	p-value
mRNA metabolic process (GO:0016071)	625	23	2.01	11.4	1.17E-14
RNA processing (GO:0006396)	930	24	2.99	8.02	4.90E-12
RNA catabolic process (GO:0006401)	235	14	0.76	18.5	2.52E-10
translation (GO:0006412)	514	18	1.65	10.9	2.83E-10
peptide biosynthetic process (GO:0043043)	539	18	1.74	10.4	6.29E-10
mRNA catabolic process (GO:0006402)	206	13	0.66	19.6	1.11E-09
viral transcription (GO:0019083)	119	11	0.38	28.7	1.69E-09
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	119	11	0.38	28.7	1.69E-09
selenocysteine metabolic process (GO:0016259)	88	10	0.28	35.3	3.00E-09
viral gene expression (GO:0019080)	131	11	0.42	26.1	4.71E-09

N.B. Ref = the number of proteins with a given annotation from the entire human reference list used (20814 entries), Obs = number of proteins observed with specified annotation, Exp = number of proteins expected by chance alone based on the given sample size, FE = fold enrichment (i.e. Obs/Exp).

Table 4.13B: Top ten enriched molecular function GO terms, by Bonferroni adjusted p-value (HeLa plus arsenite samples)

GO molecular function complete	Ref	Obs	Exp	FE	p-value
RNA binding (GO:0003723)	1718	45	5.53	8.14	2.05E-29
poly(A) RNA binding (GO:0044822)	1165	39	3.75	10.4	4.90E-28
nucleic acid binding (GO:0003676)	4514	51	14.53	3.51	9.33E-18
heterocyclic compound binding (GO:1901363)	6512	53	20.96	2.53	2.82E-12
organic cyclic compound binding (GO:0097159)	6581	53	21.18	2.5	4.62E-12
structural constituent of ribosome (GO:0003735)	231	11	0.74	14.79	6.19E-07
protein binding (GO:0005515)	10761	58	34.64	1.67	4.76E-06
binding (GO:0005488)	14353	65	46.2	1.41	1.99E-05
mRNA binding (GO:0003729)	160	8	0.52	15.53	1.44E-04
structural molecule activity (GO:0005198)	775	14	2.49	5.61	4.04E-04

N.B. Ref = the number of proteins with a given annotation from the entire human reference list used (20814 entries), Obs = number of proteins observed with specified annotation, Exp = number of proteins expected by chance alone based on the given sample size, FE = fold enrichment (i.e. Obs/Exp).

Table 4.13C: Top ten enriched cellular component GO terms, by Bonferroni adjusted p-value (HeLa plus arsenite samples)

GO cellular component complete	Ref	Obs	Exp	FE	p-value
ribonucleoprotein complex (GO:1990904)	750	29	2.41	12	3.70E-21
intracellular ribonucleoprotein complex (GO:0030529)	750	29	2.41	12	3.70E-21
macromolecular complex (GO:0032991)	4651	47	14.97	3.14	1.39E-13
ribosome (GO:0005840)	257	13	0.83	15.7	2.73E-09
cytosolic ribosome (GO:0022626)	110	10	0.35	28.2	4.14E-09
extracellular exosome (GO:0070062)	2725	32	8.77	3.65	8.41E-09
extracellular vesicle (GO:1903561)	2738	32	8.81	3.63	9.56E-09
ribosomal subunit (GO:0044391)	167	11	0.54	20.5	9.74E-09
extracellular organelle (GO:0043230)	2740	32	8.82	3.63	9.75E-09
small ribosomal subunit (GO:0015935)	70	8	0.23	35.5	1.16E-07

N.B. Ref = the number of proteins with a given annotation from the entire human reference list used (20814 entries), Obs = number of proteins observed with specified annotation, Exp = number of proteins expected by chance alone based on the given sample size, FE = fold enrichment (i.e. Obs/Exp).

4.9.8 Consideration of GO term results for HeLa plus arsenite

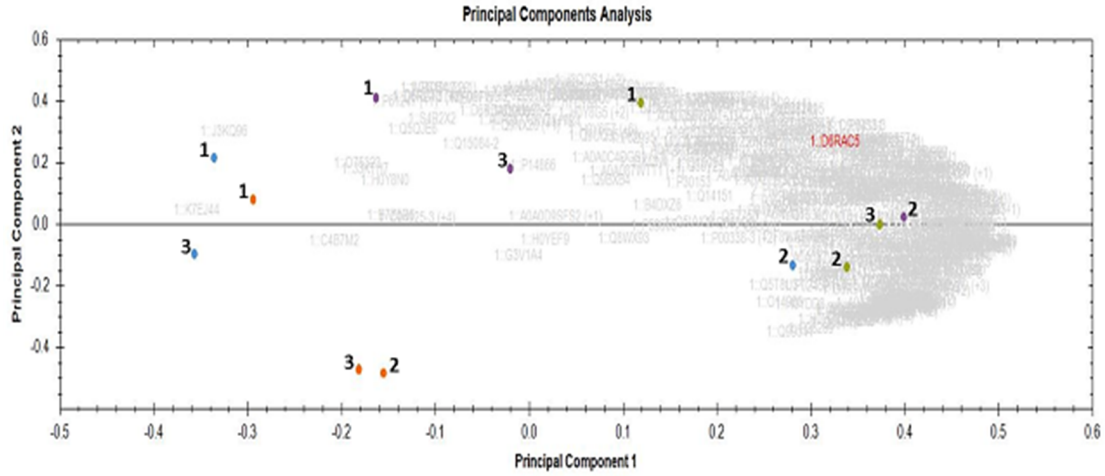
As with the minus arsenite samples, the top hits for HeLa plus arsenite were also largely related to RNA (Tables 4.13A, 4.13B, 4.13C). As there were 67 proteins in the plus arsenite dataset that were tested for GO term enrichment, as opposed to 34 for the minus arsenite samples, the p-values were much lower. It would therefore seem even more reasonable to conclude from this data set that UNR is likely to be intimately involved with RNA than was the case for the minus arsenite samples. It is unclear as to why more RNA-associated proteins were considered significant in the plus arsenite samples than in the minus arsenite samples. A working hypothesis in the lab is that UNR is localised to stress granules following arsenite stress which may make it easier to pull down. Assuming that the immunoprecipitating antibody is saturated, that would make UNR from a pool bound to non-RNA-associated proteins become diluted in the end sample. Alternatively, it could allow the antibody to become more saturated with the possibly more accessible stress granule-located UNR.

Viral transcription was considered among the most significant biological processes again, as were selenocysteine metabolic processes (Table 4.13A) and extracellular exosome/vesicle/organelle (Table 4.13C).

4.10 Progenesis results for U2OS RIP dataset

The U2OS samples were initially imported into Progenesis with 348082 MS/MS spectra, of which 50393 were re-imported from MASCOT having been assigned to peptide ions. As with the HeLa samples, expected contaminants were removed from the data. As before, proteins with an ANOVA p-value over 0.05 were also removed as this caused the samples to group more closely by treatment over PC1 and PC2 (Figure 4.15). This means that the main sources of variability in the data were more closely aligned with the treatments (arsenite treatment and immunoprecipitating immunoglobulin). Generally positive normalised and standardised values for PC1 (above 0.1) are UNR pulldown samples. Negative values (strictly, less than 0.1) are IgG pulldown samples. This implies that the main source of variation in the data (accounting for 57.0% of all the variation) is accounted for by immunoprecipitating antibody (Figure 4.15B). Likewise, positive values in PC2 are linked to plus arsenite samples and negative PC2 samples (other than one slightly negative value that touches the line PC2=0) are unstressed samples. This means that arsenite treatment accounts for most of the second largest component of variability in the data, which itself accounts for 23.4% of the total variation (Table 4.14B).

A



B

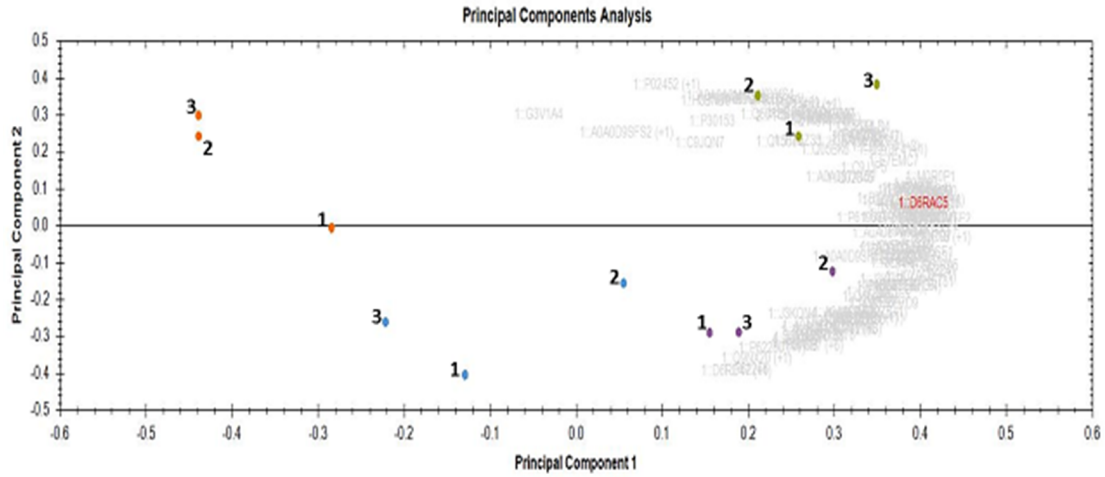


Figure 4.15: Progenesis-generated PCA plots for U2OS data. Proteins with all ANOVA p-values (A) or only those with ANOVA p-values under 0.05 (B) were included. See text for further information how the data were processed. Numbers are associated with the days upon which the lysates were made (i.e. the repeats). Blue = no arsenite, IgG; purple = no arsenite, UNR; orange = plus arsenite, IgG; green = plus arsenite, UNR. In terms of variance accounted for – A: (PC1 = 49.42%, PC2 = 19.89%); B: (PC1 = 56.98%, PC2 = 23.36%).

Having exported the data into Excel, paired t-tests were carried out on the plus and minus arsenite U2OS samples separately. This suggested 63 putative UNR interacting proteins from the minus arsenite samples when the cut off p-value was set at 0.1 (Table 4.14B). The top ten putative UNR-interactors by p-value are presented below (Table 4.14A). Meanwhile, 87 putative

hits were suggested from the U2OS plus arsenite data (Table 4.15B). The top ten values are additionally presented below (Table 4.15A).

Table 4.14A: Top 10 putative U2OS minus arsenite UNR interacting proteins by t-test p-value (using Progenesis data)

protein	U2OS1	U2OS2	U2OS3	U2OS4	U2OS5	U2OS6	p-value
GUCY2D	0	3011	868	156011	153812	160214	0.00026
TRA2B	153315	286102	158026	241178	376746	241157	0.00063
SYNCRIP	181666	115723	118102	364396	306211	320994	0.00093
RPN1	254	2048	2601	13934	14081	15274	0.00140
NARR	2069	9147	6106	545559	530244	467932	0.00228
RPL35	143303	131817	124007	182065	170266	171242	0.00478
RPS8	938342	971227	461141	1526198	1553166	931819	0.00482
HNRNPR	314777	198519	237638	770567	572489	718462	0.00541
FBL	8003	21814	1442	20538	36152	12530	0.00547
HNRNPH1	50847	49523	11328	86975	76420	42451	0.00715

N.B The t-test p-values are for a two tailed paired t-test without multiple testing correction.

Table 4.14B: Gene names for putative UNR-interacting proteins from unstressed U2OS, by t-test p-value

Description	p-value	Description	p-value	Description	p-value
GN=GUCY2D	0.00026	GN=PCBP2	0.01665	GN=KPNB1	0.05098
GN=TRA2B	0.00063	GN=RPS8	0.01739	GN=RPL27	0.05193
GN=SYNCRIP	0.00093	GN=SLC25A6	0.01875	GN=RPS2	0.05556
GN=RPN1	0.00140	GN=CSDE1	0.01878	GN=RPS3	0.05655
GN=RAB34 (NARR)	0.00228	GN=RPS9	0.01958	GN=RPL29	0.05655
GN=RPL35	0.00478	GN=TRIP6	0.02137	GN=RPS3	0.05702
GN=RPS8	0.00482	GN=OAT	0.02323	GN=RPS2	0.05739
GN=HNRNPR	0.00541	GN=LDB1	0.02522	GN=FBL	0.05956
GN=FBL	0.00547	GN=RPS26P11	0.02896	GN=RPS2	0.06098
GN=HNRNPH1	0.00715	GN=RBMX	0.03079	GN=RPS2	0.06255
GN=HNRNPR	0.00797	GN=RPL26	0.03292	GN=RPS6	0.06266
GN=LMO7	0.00873	GN=DDX5	0.03305	GN=C19orf68	0.06281
GN=HNRNPH1	0.00900	GN=SSBP2	0.03373	GN=CSDE1	0.06393
GN=SSBP2	0.01091	GN=DDX5	0.03755	GN=RPS6	0.06652
GN=DDX3X	0.01326	GN=RPS9	0.03778	GN=HNRNPU	0.07406
GN=DDX5	0.01330	GN=RPL36A	0.03987	GN=HUWE1	0.08014
GN=SLC25A5	0.01544	GN=KPNB1	0.03992	GN=SRSF7	0.08554
GN=CSDE1	0.01552	GN=RPS3	0.04192	GN=CSDE1	0.09049
GN=EWSR1	0.01566	GN=HNRNPF	0.04228	GN=RPS15A	0.09157
GN=DDX5	0.01660	GN=RPS9	0.04754	GN=HNRNPK	0.09443
GN=C4A	0.01663	GN=RPS2	0.04987	GN=HNRNPA3	0.09911

N.B. Only gene names were given for the putative UNR-interacting proteins to save table space due to the large number of suggested hits. This means that some identifiers are repeated due to different isoforms being suggested. The t-test p-values are for a two tailed paired t-test without multiple testing correction and increase from top to bottom for each column, starting on the left.

Table 4.15A: Top ten putative U2OS plus arsenite UNR interacting proteins by t-test p-value (using Progenesis data)

protein	U2OS7	U2OS8	U2OS9	U2OS10	U2OS11	U2OS12	p-value
HNRNPF	46540	59224	31029	690448	674700	629441	0.00046
ANXA2	1786.9	4173.2	3312.8	32360.6	37457.4	33359.8	0.00102
HNRNPR	186734	44852	16708	382456	220243	209490	0.00114
FLNA	17574	43072	14970	380004	362924	318690	0.00283
SSBP2	70.834	107.77	33.396	1119110	1072048	913000	0.00361
HNRNPK	56620	66225	50738	929805	958167	1159729	0.00618
LDB1	28094	15251	17432	739164	914811	961052	0.00694
HNRNPR	207000	57053	24121	374470	264043	250067	0.00733
C4A	5576.5	10878	12473	163916	138351	131491	0.00775
FLNC	30001	53146	40043	646044	486618	613826	0.01024

N.B. The t-test p-values are for a two tailed paired t-test without multiple testing correction.

Table 4.15B: Gene names for putative UNR-interacting proteins from arsenite-treated U2OS, by t-test p-value

Description	p-value	Description	p-value	Description	p-value
GN=HNRNPF	0.00046	GN=GUCY2D	0.02816	GN=RPS6	0.05906
GN=ANXA2	0.00102	GN=FASN	0.02861	GN=RPS3	0.05989
GN=HNRNPR	0.00114	GN=RBMX	0.02897	GN=STRAP	0.06014
GN=FLNA	0.00283	GN=RPS2	0.03100	GN=RPS3	0.06639
GN=SSBP2	0.00361	GN=DDX5	0.03141	GN=MRPS7	0.06905
GN=HNRNPK	0.00618	GN=C19orf68	0.03375	GN=RPS6	0.06943
GN=LDB1	0.00694	GN=UQCRQ	0.03400	GN=SLC25A6	0.07148
GN=HNRNPR	0.00733	GN=RPS2	0.03405	GN=SLC25A5	0.07250
GN=C4A	0.00775	GN=HNRNPU	0.03443	GN=RPS16	0.07277
GN=FLNC	0.01024	GN=RPS13	0.03458	GN=RPS11	0.07285
GN=HUWE1	0.01235	GN=RPS2	0.03474	GN=CSDE1	0.07301
GN=SYNCRIP	0.01265	GN=RPS8	0.03662	GN=RPS11	0.07560
GN=SSBP2	0.01325	GN=RPS2	0.03737	GN=PCBP1	0.07698
GN=SRSF2	0.01365	GN=DARS	0.03768	GN=VIM	0.07699
GN=HNRNPH1	0.01559	GN=RPS2	0.03779	GN=MRPL16	0.07838
GN=HNRNPK	0.01673	GN=RPS8	0.03823	GN=DDX3X	0.07874
GN=FLNA	0.01788	GN=RPL23	0.04232	GN=EIF3B	0.07904
GN=C1orf167	0.01819	GN=DDX5	0.04656	GN=HNRNPH1	0.08205
GN=FBL	0.02070	GN=TRIM28	0.04732	GN=TRA2B	0.08516
GN=OAT	0.02190	GN=PTRF	0.04734	GN=RPS9	0.08677
GN=DDX5	0.02208	GN=RPS14	0.04770	GN=HNRNPA2B1	0.08780
GN=RPS13	0.02289	GN=SQSTM1	0.04778	GN=RPS3	0.08941
GN=RPS4X	0.02401	GN=DDX5	0.04778	GN=RBMX	0.09033
GN=RPS10-NUDT3	0.02429	GN=PCBP2	0.04798	GN=CSDE1	0.09342
GN=RPS15A	0.02458	GN=HSPA1A	0.04814	GN=RPS3	0.09586
GN=RPL36A	0.02485	GN=RPS9	0.04887	GN=RPS23	0.09626
GN=RPS26P11	0.02518	GN=RPS3	0.05618	GN=TFG	0.09767
GN=SRSF3	0.02534	GN=RAB34 (NARR)	0.05740	GN=TAF15	0.09863
GN=LMO7	0.02770	GN=SQSTM1	0.05806	GN=PCBP2	0.09998

N.B. Only gene names were given for the putative UNR-interacting proteins to save table space due to the large number of suggested hits. This means that some identifiers are repeated due to different isoforms being suggested. The t-test p-values are for a two tailed paired t-test without multiple testing correction and increase from top to bottom for each column, starting on the left.

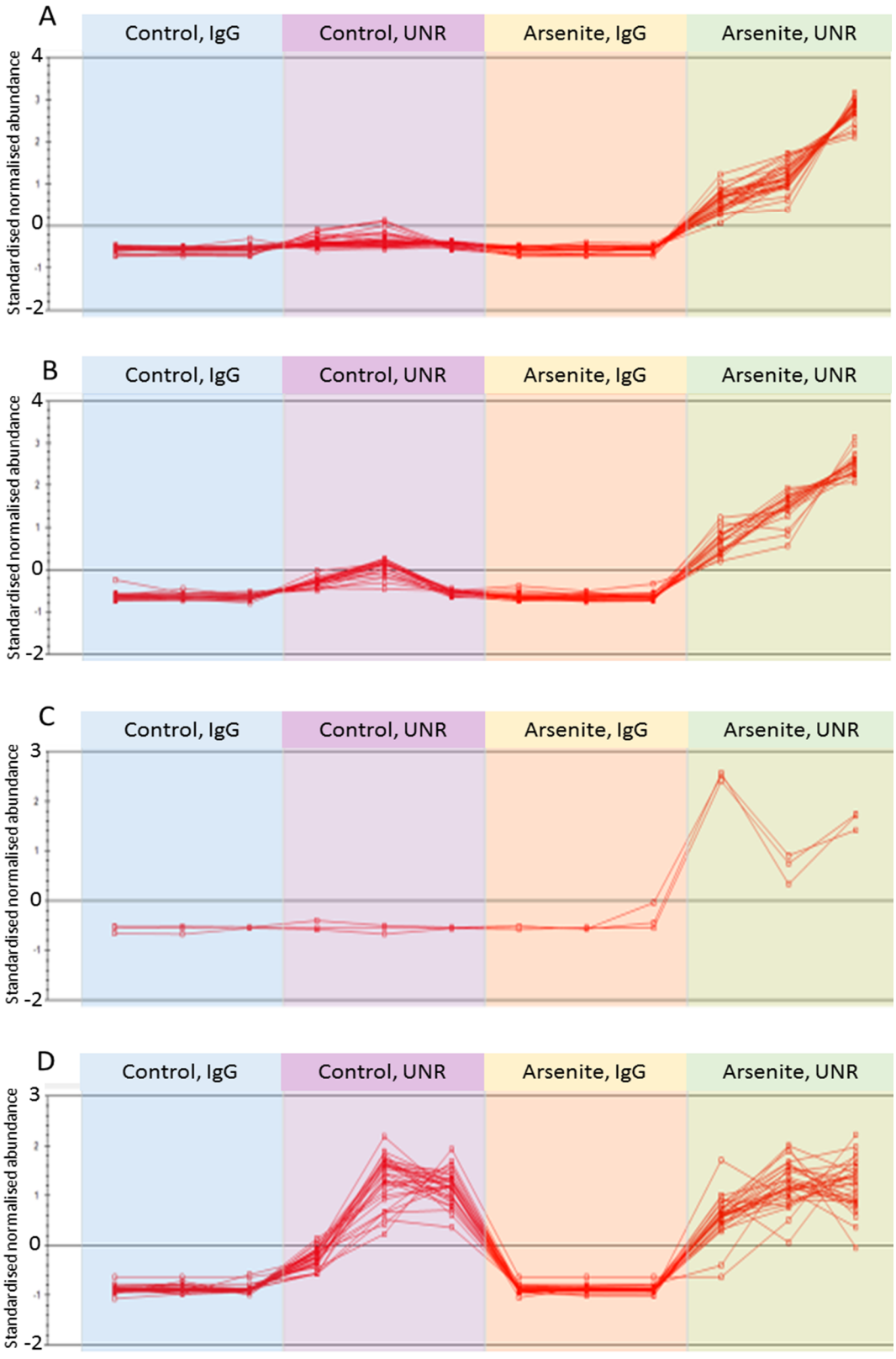


Figure 4.16: Standardised normalised abundances for peptides assigned to specific proteins (A=UNR ('different' forms combined), B = UNRIP, C = TRIM28, D = HUWE1) from the U2OS data. The repeats in the blue column are IgG (minus arsenite), the repeats in the purple column are UNR (minus arsenite), the repeats in the orange column are IgG (plus arsenite) and the repeats in the green column are UNR (plus arsenite).

On this occasion, as well as UNR and UNRIP, it was decided to show peptide data for the TRIM33-related TRIM28 (also known as TIF1-β) and HUWE1. Each of these proteins shows something about the data. Firstly, it must be assumed that UNR is a genuine hit; should it not be a true hit in both the plus and minus arsenite samples, then the entire experiment and all inferred results are invalid. It can be seen that much more UNR was detected in the plus arsenite UNR pulldowns (Figure 4.16A, green column) than in the minus arsenite samples (Figure 4.16A, purple column), although both are higher than the IgG samples in the other columns. It can also be seen that the peptides that were assigned to UNR all followed the same basic detection level pattern between the different samples (Figure 4.16A). This implies that they were correctly assigned. A very similar pattern was seen for UNRIP as was seen for UNR, given that the scale on the y axis is the same between the two proteins – ranging between -2 and +4 (Figures 4.16B, 4.16A). This could imply that the levels of UNR and UNRIP vary together and both increase upon arsenite stress. It could also be that UNR and UNRIP interact together in unstressed cells and stressed cells equally but that the RIP process becomes more efficient in stressed cells, possibly due to UNR and UNRIP becoming concentrated in stress granules. TRIM28 shows an interesting peptide pattern – only three peptides were assigned to it but the detected levels of each are tightly correlated with the samples, implying that the assignment was probably correct. TRIM28 appears to only be associated with UNR in arsenite stressed cells (Figure 4.16C). Most, but not all, the peptides assigned to the last protein, HUWE1, follow a constant pattern between the samples (Figure 4.16D). This implies that some (3 or 4) of the peptides may have been misassigned. HUWE1 appears to associate strongly with UNR in U2OS whether or not the cells have been stressed (Figure 4.16D, purple/green UNR pulldown columns versus blue/orange IgG pulldown columns).

Finally, a fold change minimum of 10 was set as a significance cut off point (again only considering the data with ANOVA p-values of 0.05 or less). This gave a set of ten putative UNR-interactors for the minus arsenite U2OS samples (Table 4.16) and 48 for the plus arsenite U2OS samples (Table 4.17B). The top ten hits by ratio for U2OS plus arsenite are presented in Table 4.17A.

Table 4.16: Putative U2OS minus arsenite UNR interacting proteins by ratio (using Progenesis data)

Protein	U2OS1	U2OS2	U2OS3	U2OS4	U2OS5	U2OS6	ratio
SSBP2 (A0A087X159)	0	0	0	274510	546075	433402	∞
SSBP2 (D6RAC5)	108	285	153	767310	842104	1080423	4932
TFG (Q05BK6)	10	1	33	474	5451	58232	1445
GUCY2D	0	3011	868	156011	153812	160214	121
NARR	2069	9147	6106	545559	530244	467932	89
LMO4	7194	0	171	52845	79396	383006	70
OAT	24	892	0	10631	18960	13406	47
HUWE1	96423	138564	73000	1299253	4195472	3848769	30
LDB1	29913	35416	15096	591143	994118	688242	28
TFG (Q05BK6)	194	18301	0	6502	56466	161279	12

N.B. Ratios are mean(UNR)/mean(IgG) and all infinities were ignored unless every UNR value was greater than zero. The yellow shading highlights a protein absent in every IgG sample but present in every UNR sample.

Table 4.17A: Top 10 putative U2OS plus arsenite UNR interacting proteins by ratio (using Progenesis data)

Protein	U2OS7	U2OS8	U2OS9	U2OS10	U2OS11	U2OS12	ratio
PPIA	0	0	0	53820	6285	13257	∞
PABPC5	0	0	0	3312	5290	87639	∞
SSBP2 (D6RAC5)	71	108	33	1119110	1072048	913000	14642
PDIA6	0	0	10	82803	1166	1551	8868
SQSTM1	12	3	0	12442	17093	26903	3749
TFG (C9JJP5)	0	25	55	22380	79601	95845	2464
LMO4	0	0	593	117167	159453	589844	1461
OAT	180	0	0	36907	27848	21809	480
RANGAP1	176	275	0	19614	7132	144834	381
SSBP2 (A0A087X159)	2675	1646	0	407577	587656	591665	367

N.B. Ratios are mean(UNR)/mean(IgG) and all infinities were ignored unless every UNR value was greater than zero. The yellow shading highlights a protein absent in every IgG sample but present in every UNR sample.

Table 4.17B: Putative U2OS plus arsenite UNR interacting proteins by mean UNR/IgG ratios (using Progenesis data)

Description	ratio	Description	ratio	Description	ratio
GN=PPIA	∞	GN=ATP5O	57	GN=RNH1	17
GN=PABPC5	∞	GN=CSDE1	51	GN=C4A	15
GN=SSBP2	14642	GN=CFL1	51	GN=HNRNPF	15
GN=PDIA6	8868	GN=RANBP2	51	GN=EIF2A	14
GN=SQSTM1	3749	GN=LDB1	43	GN=FLNC	14
GN=TFG	2464	GN=FASN	39	GN=FLNA	14
GN=LMO4	1461	GN=PCBP2	38	GN=VIM	14
GN=OAT	480	GN=RPN1	37	GN=PPP2R1A	14
GN=RANGAP1	381	GN=CSDE1	35	GN=RPS3	13
GN=SSBP2	367	GN=HUWE1	33	GN=FBL	13
GN=GUCY2D	320	GN=FHL2	33	GN=RBM14	13
GN=RAB34 (NARR)	170	GN=STRAP	32	GN=SQSTM1	12
GN=TFG	118	GN=KPNB1	30	GN=PCBP2	12
GN=EIF3B	86	GN=TRIM28	22	GN=HSPA1A	12
GN=CSDE1	68	GN=KPNB1	18	GN=HNRNPK	11
GN=CSDE1	59	GN=HNRNPK	18	GN=ANXA2	11

N.B. Only gene names were given for the putative UNR-interacting proteins to save table space due to the large number of suggested hits. This means that some identifiers are repeated due to different isoforms being suggested. The t-test p-values are for a two tailed paired t-test without multiple testing correction and increase from top to bottom for each column, starting on the left.

4.10.1 Discussion of U2OS RIP dataset in comparison to the HeLa dataset

The top ten putative UNR interactors by t-test p-value for the U2OS minus and plus arsenite (Tables 4.14A, 4.15A) were quite different to those for HeLa minus and plus arsenite (Tables 4.7A, 4.8A). In terms of the top ten putative interactors by fold enrichment in UNR pulldowns over IgG pulldowns, there was more similarity between the cell types. SSBP2 (twice in U2OS), NARR, LMO4, HUWE1 and LBD1 were all present in the top ten HeLa minus arsenite list (Table 4.9A) and the U2OS minus arsenite list (Table 4.16). Some of those proteins had very large fold enrichments in the UNR samples. SSBP2, for example, had a UNR/IgG ratio of 276 in HeLa minus

arsenite and the two entries for U2OS minus arsenite had ratios of almost 5000 and infinity, respectively.

The top ten by fold enrichment for U2OS plus arsenite (Table 4.17A) and HeLa plus arsenite (Table 4.10A) also had some cross-over. SSBP2 (twice in U2OS), TFG, LMO4 and RANGAP1 were on both lists. The lowest entry on the top ten fold enrichment list for U2OS plus arsenite had a rounded UNR/IgG ratio of 367:1 whereas it the tenth most significant entry on the HeLa plus arsenite ratio list only had a UNR/IgG fold increase of 57.

Another point of discussion was the number of hits between the two cell types. There were 25 proteins with t-test p-values less than 0.1 for HeLa minus arsenite. The number for U2OS was 64. By fold enrichment, HeLa minus arsenite had 17 proteins with fold enrichments over 10, in U2OS, there were only 10. With the plus arsenite samples, HeLa had 67 proteins with a p-value less than 0.1, whereas U2OS had 88. By fold enrichment, there were 17 proteins with a fold enrichment over 10, whereas the number was 48 for U2OS.

4.10.2 Progenesis results for U2OS RIP dataset exported to AmiGO 2

The plus and minus arsenite hits suggested by Progenesis were exported separately to AmiGO 2 version 2.3.2 following removal of NARR, TRIP6 (for minus arsenite as was higher minus arsenite IgG compared to minus arsenite UNR and was only present as it was highest in plus arsenite UNR overall) and heavy and light chains (for both plus and minus arsenite) that were assigned to non-IgG antibodies. Other than that, processing was as per the HeLa samples (see section 4.9.4).

4.10.3 GO term results for U2OS minus arsenite

A total of 45 individual proteins were entered into AmiGO 2 and all of them were recognised by the tool.

A Bonferroni multiple testing corrected p-value of 0.05 was chosen as the significance cut off point and top ten lists of overrepresented GO terms by p-value, for each set of GO terms, were recorded in Table 4.18 (Table 4.18A = biological process, Table 4.18B = molecular function, Table 4.18C = cellular component).

Table 4.18A: Top ten enriched biological process GO terms, by Bonferroni adjusted p-value (U2OS minus arsenite samples)

GO biological process complete	Ref	Obs	Exp	FE	p-value
mRNA metabolic process (GO:0016071)	625	24	1.35	17.76	4.87E-21
RNA processing (GO:0006396)	930	24	2.01	11.94	4.97E-17
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	111	12	0.24	50	1.07E-13
cotranslational protein targeting to membrane (GO:0006613)	114	12	0.25	48.69	1.46E-13
protein targeting to ER (GO:0045047)	114	12	0.25	48.69	1.46E-13
establishment of protein localization to endoplasmic reticulum (GO:0072599)	118	12	0.26	47.04	2.20E-13
selenocysteine metabolic process (GO:0016259)	88	11	0.19	57.82	5.66E-13
protein localization to endoplasmic reticulum (GO:0070972)	138	12	0.3	40.22	1.40E-12
selenium compound metabolic process (GO:0001887)	114	11	0.25	44.63	9.39E-12
viral transcription (GO:0019083)	119	11	0.26	42.76	1.49E-11

Table 4.18B: Top ten enriched molecular function GO terms, by Bonferroni adjusted p-value (U2OS minus arsenite samples)

GO molecular function complete	Ref	Obs	Exp	FE	p-value
poly(A) RNA binding (GO:0044822)	1165	33	2.52	13.1	1.92E-28
RNA binding (GO:0003723)	1718	34	3.71	9.15	1.61E-24
nucleic acid binding (GO:0003676)	4514	37	9.76	3.79	2.41E-14
structural constituent of ribosome (GO:0003735)	231	14	0.5	28.03	1.41E-13
heterocyclic compound binding (GO:1901363)	6512	39	14.08	2.77	5.19E-11
organic cyclic compound binding (GO:0097159)	6581	39	14.23	2.74	7.62E-11
mRNA binding (GO:0003729)	160	11	0.35	31.8	1.20E-10
structural molecule activity (GO:0005198)	775	14	1.68	8.36	1.49E-06
binding (GO:0005488)	14353	45	31.03	1.45	1.48E-04
single-stranded RNA binding (GO:0003727)	65	5	0.14	35.58	8.87E-04

Table 4.18C: Top ten enriched cellular component GO terms, by Bonferroni adjusted p-value (U2OS minus arsenite samples)

GO cellular component complete	Ref	Obs	Exp	FE	p-value
ribonucleoprotein complex (GO:1990904)	750	24	1.62	14.8	5.34E-20
intracellular ribonucleoprotein complex (GO:0030529)	750	24	1.62	14.8	5.34E-20
cytosolic ribosome (GO:0022626)	110	13	0.24	54.66	2.02E-16
ribosomal subunit (GO:0044391)	167	13	0.36	36.01	4.24E-14
cytosolic part (GO:0044445)	225	14	0.49	28.78	4.68E-14
ribosome (GO:0005840)	257	13	0.56	23.4	1.01E-11
cytosolic small ribosomal subunit (GO:0022627)	41	8	0.09	90.25	5.91E-11
macromolecular complex (GO:0032991)	4651	34	10.06	3.38	6.64E-11
catalytic step 2 spliceosome (GO:0071013)	88	9	0.19	47.3	4.31E-10
small ribosomal subunit (GO:0015935)	70	8	0.15	52.86	4.08E-09

4.10.4 Consideration of GO term results for U2OS minus arsenite

The top ten enriched biological process GO terms contained were quite varied (Table 4.18A). RNA-related proteins were the most numerous in terms of the number of observed proteins. Other top hits included terms related to protein localisation, viral transcription and selenium metabolism. The top molecular function GO terms (Table 4.18B) implied a previously established role for UNR in translation with the top hit being poly(A) RNA binding and other top hits including mRNA binding, structural component of ribosome and single stranded RNA binding. The most enriched cellular component GO terms were also largely related to RNA processes. The top hit was ribonucleoprotein complex and other top hits included ribosome and catalytic step 2 spliceosome. These results point to a clear role for UNR in the translation and related processes. The observation that 12.5% of all proteins annotated to selenocysteine metabolic process (11 out of 88) is interesting. Whilst it is not going to be considered further here, the information is there for workers in the field of selenocysteine biochemistry who may wish to explore any potential role for UNR in relation to their research.

4.10.5 GO term results for U2OS plus arsenite

A total of 76 individual proteins were entered into AmiGO 2 and all of them were recognised by the tool. It should be noted that, due to inconsistencies in the proteins used for the initial GO term analysis for U2OS plus arsenite (an immunoglobulin had remained in the putative UNR-interactor data), it was decided to run the analysis again with an updated database (Table 4.19A). There was no change in the identity of the top GO terms and little change in their order. The number of proteins annotated to many GO terms in the reference set had changed, however, and this resulted in small changes to the p-values. It is not expected that this should change any inferences about UNR-interactors but any direct comparisons with the U2OS minus arsenite results should take this difference into consideration.

A Bonferroni multiple testing corrected p-value of 0.05 was chosen as the significance cut off point and the top ten UNR-enriched GO terms were recorded in Table 4.19 (Table 4.19B = biological process, Table 4.19C = molecular function, Table 4.19D = cellular component).

Table 4.19A: Technical data pertaining to the GO term enrichment analyses carried out on putative UNR interacting proteins from arsenite treated U2OS cells

Analysis Type:	PANTHER Overrepresentation Test (release 20160715)
Annotation Version and Release Date:	GO Ontology database Released 2016-07-29
Analyzed List:	upload_1 (Homo sapiens)
Reference List:	Homo sapiens (all genes in database)
Bonferroni correction:	TRUE

Table 4.19B: Top ten enriched biological process GO terms, by Bonferroni adjusted p-value (U2OS plus arsenite samples)

GO biological process complete	Ref	Obs	Exp	FE	p-value
mRNA metabolic process (GO:0016071)	621	32	2.28	14.03	1.24E-24
RNA processing (GO:0006396)	860	31	3.16	9.82	4.05E-19
multi-organism metabolic process (GO:0044033)	138	18	0.51	35.53	4.93E-19
viral gene expression (GO:0019080)	126	17	0.46	36.75	4.92E-18
viral life cycle (GO:0019058)	292	21	1.07	19.59	1.61E-17
translational initiation (GO:0006413)	153	17	0.56	30.26	1.24E-16
translation (GO:0006412)	450	23	1.65	13.92	2.73E-16
viral process (GO:0016032)	822	28	3.02	9.28	3.94E-16
multi-organism cellular process (GO:0044764)	825	28	3.03	9.24	4.33E-16
cellular nitrogen compound metabolic process (GO:0034641)	5131	57	18.84	3.03	6.94E-16

Table 4.19C: Top ten enriched molecular function GO terms, by Bonferroni adjusted p-value (U2OS plus arsenite samples)

GO molecular function complete	Ref	Obs	Exp	FE	p-value
RNA binding (GO:0003723)	1623	53	5.96	8.89	2.72E-37
poly(A) RNA binding (GO:0044822)	1163	47	4.27	11.01	9.60E-36
nucleic acid binding (GO:0003676)	4046	56	14.86	3.77	1.22E-20
structural constituent of ribosome (GO:0003735)	226	20	0.83	24.1	9.57E-19
heterocyclic compound binding (GO:1901363)	5922	59	21.74	2.71	4.91E-15
organic cyclic compound binding (GO:0097159)	6005	59	22.05	2.68	1.01E-14
mRNA binding (GO:0003729)	164	15	0.6	24.91	1.47E-13
rRNA binding (GO:0019843)	58	10	0.21	46.96	6.36E-11
structural molecule activity (GO:0005198)	774	22	2.84	7.74	1.11E-10
protein binding (GO:0005515)	10839	68	39.8	1.71	2.26E-08

Table 4.19D: Top ten enriched cellular component GO terms, by Bonferroni adjusted p-value (U2OS plus arsenite samples)

GO cellular component complete	Ref	Obs	Exp	FE	p-value
ribonucleoprotein complex (GO:1990904)	762	35	2.8	12.51	1.14E-26
intracellular ribonucleoprotein complex (GO:0030529)	762	35	2.8	12.51	1.14E-26
cytosolic small ribosomal subunit (GO:0022627)	45	15	0.17	90.79	3.71E-22
small ribosomal subunit (GO:0015935)	70	16	0.26	62.25	3.39E-21
ribosomal subunit (GO:0044391)	172	19	0.63	30.09	9.45E-20
cytosolic ribosome (GO:0022626)	121	17	0.44	38.27	3.90E-19
macromolecular complex (GO:0032991)	4795	57	17.61	3.24	3.42E-18
ribosome (GO:0005840)	250	19	0.92	20.7	9.36E-17
adherens junction (GO:0005912)	680	26	2.5	10.41	1.12E-16
anchoring junction (GO:0070161)	698	26	2.56	10.15	2.12E-16

4.10.6 Consideration of GO term results for U2OS plus arsenite

The most enriched biological process GO terms included a number that were related to RNA processing and translational initiation (Table 4.19B). Interestingly, virus-related GO terms were among the most numerous enriched GO terms, both in terms of the maximum number of proteins annotated to a particular GO term (28 to viral processes was the third highest after the 32 to mRNA metabolic processes and 31 to the highly related RNA processing GO term) and also in terms of the number of GO terms – 4 out of the top 10.

When it is considered that these GO terms become top hits in stressed U2OS cells, it possibly implies a role for UNR in both viral infection and cancer. It is known that Unr is among other pieces of proteinaceous cellular machinery that can become usurped by viruses following infection (Boussadia *et al.*, 2003). Hijacking or neutralisation of cellular proteins is a common theme in both viral infection and the development of the cancer phenotype. For example, the DNA virus adenovirus reprograms infected cells to create a pseudo-S phase environment that is conducive to viral replication using a combination of virally-expressed and hijacked cellular proteins (Turnell, 2008). By generating an S phase-like environment in which DNA can be replicated without appropriate growth signals, adenovirus-infected cells exhibit one of Hanahan and Weinberg's six classic hallmarks of cancer (Hanahan & Weinberg, 2000). Whilst an in-depth discussion of the molecular biology of adenovirus is beyond the scope of this work (for review, see Leppard, (2014)), it should be considered that its E1 region has transformed a variety of cell lines, starting with HEK293 (Graham *et al.*, 1977), thereby giving the cells limitless replicative potential – another classic hallmark of cancer (Hanahan & Weinberg, 2000). The large E1B protein, in conjunction with the E4orf6 protein, can target TP53 for degradation – helping the virally infected cell to overcome antigrowth signals and apoptosis (Turnell, 2008). Other evidence shows that adenovirus can also help a cell to evade immune detection, by reducing the cell surface expression of both major histocompatibility complex (MHC) class I proteins and activating natural killer cell receptors (Sester et al. 2010). That prevents detection of cell infection by both T cells and natural killer cells and thereby confers one of Hanahan and Weinberg's 'emerging' hallmarks of cancer upon the infected cell (Hanahan & Weinberg, 2011). To summarise, a link between UNR and viral life cycles could well provide insight into any potential role of UNR in cancer because, as has been briefly considered here in relation to adenovirus, similar mechanisms are involved in both. The important point in relation to UNR is that it interacts with a large number of proteins that may be targeted by viruses as part of their

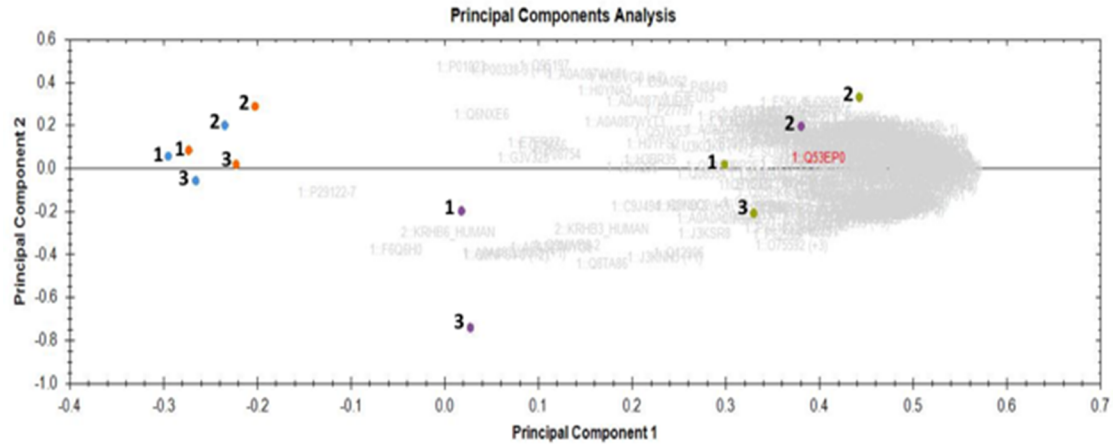
life cycle. In such, certain viruses may have a selective pressure to impede one or more physiological functions of UNR, as may certain cancer cells.

As with the U2OS minus arsenite samples, most of the top ten enriched molecular function GO terms (Table 4.19C) and cellular component GO terms (Table 4.19D) were related to RNA and translation. Interestingly, however, two of the top ten cellular component GO terms were to do with cell-cell interaction, e.g. adherens junction. It was later shown that UNR (and HUWE1), are found concentrated at areas close to points of cell-cell interaction in unstressed U2OS cells (Figure 4.22) and in stressed HeLa cells that appeared to be undergoing cytokinesis (Figure 4.23).

4.11 Progenesis results for SaOS-2 RIP dataset

The U2OS samples were initially imported into Progenesis with 319420 MS/MS spectra, of which 27819 were re-imported from MASCOT having been assigned to peptide ions. As with the HeLa and U2OS samples, it was decided to remove proteins with an ANOVA p-value over 0.05 (Figure 4.17).

A



B

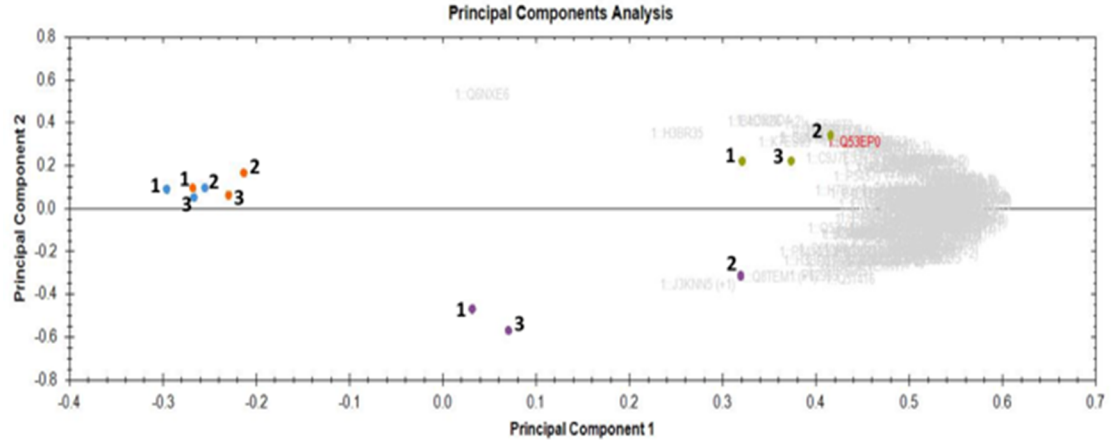


Figure 4.17: Progenesis-generated PCA plots for SaOS-2 data. Proteins with all ANOVA p-values (A) or only those with ANOVA p-values under 0.05 (B) were included. See text for further information how the data were processed. Numbers are associated with the days upon which the lysates were made (i.e. the repeats). Blue = no arsenite, IgG; purple = no arsenite, UNR; orange = plus arsenite, IgG; green = plus arsenite, UNR. In terms of variance accounted for – A: (PC1 = 66.03%, PC2 = 9.07%); B: (PC1 = 77.72%, PC2 = 7.64%).

The PC1vPC2 PCA image in Figure 4.17(B) shows that the line PC1 = 0 delineates the IgG samples (PC1 < 0) from the UNR samples (PC1 > 0). The line PC2 = 0 also separates the minus arsenite UNR samples (PC2 < 0) plus arsenite UNR samples (PC2 > 0), PC2 does not distinguish between the plus and minus arsenite IgG samples. That could be explained by arsenite treatment only changing the SaOS-2 proteome by a small amount, but the changes that did occur being UNR-specific. That way, the background available to bind to IgG would have been fairly similar,

leading to clustering of the plus and minus arsenite IgG samples. The UNR samples, however, could have changed more substantially, leading to clear separation of the plus and minus arsenite UNR samples by PC2 (Figure 4.17B). It should also be considered that principal components do not have to be releated exclusively to a given treatment and are likely to be influenced by different treatments and, more so with lower components, background noise.

Having exported the data into Excel, paired t-tests were carried out on the plus and minus arsenite SaOS-2 samples separately. This suggested 39 putative UNR interacting proteins from the minus arsenite samples when the cut off p-value was set at 0.1 (Table 4.20B). The top ten suggested proteins are presented in Table 4.20A. Meanwhile, 176 putative hits were suggested for the plus arsenite data (Table S1). The top ten of these are tabulated below (Table 4.21).

Table 4.20A: Top 10 putative SaOS-2 minus arsenite UNR interacting proteins by t-test p-value (using Progenesis data)

Description	SaOS1	SaOS2	SaOS3	SaOS4	SaOS5	SaOS6	p-value
SQSTM1 (E3W990)	1129	629	2736	469171	438770	432590	0.00068
TBC1D10C	25531	20939	15411	118367	117525	103569	0.00069
CCDC50	8485	11291	9414	437035	427868	399030	0.00078
LDB1	79228	62590	91126	6597521	7479055	6860085	0.00150
ATRX	16515	8301	15982	137541	115317	116785	0.00296
SQSTM1 (E7EMC7)	1888	4624	4903	492722	601250	610768	0.00425
FXR1	26395	17312	17945	15140	7862	9106	0.00538
SSBP3	41088	27850	95677	9575659	11932766	9774793	0.00543
C4A	866	576	1198	18148	16042	21293	0.00579
LMO7	1072254	833296	882401	2227386	2185545	1915409	0.00615

N.B. The t-test p-values are for a two tailed paired t-test without multiple testing correction.

Table 4.20B: Gene names for putative SaOS-2 minus arsenite UNR interacting proteins by t-test p-value (using Progenesis data)

Description	p-value	Description	p-value	Description	p-value
GN=SQSTM1	0.00068	GN=LDB2	0.01258	GN=UBR5	0.05733
GN=TBC1D10C	0.00069	GN=SQSTM1	0.01360	GN=ALK	0.05920
GN=CCDC50	0.00078	GN=TRMT1L	0.01781	GN=RPS26P11	0.06206
GN=LDB1	0.00150	GN=SIRT1	0.01847	GN=ERCC4	0.06351
GN=ATRX	0.00296	GN=ARHGEF2	0.01990	GN=H3F3A	0.07115
GN=SQSTM1	0.00425	GN=RPS4X	0.02033	GN=RPS25	0.07260
GN=FXR1	0.00538	GN=RYR1	0.02089	GN=HUWE1	0.07327
GN=SSBP3	0.00543	GN=RAB34 (NARR)	0.02114	GN=ATP5F1	0.07865
GN=C4A	0.00579	GN=CCDC137	0.02986	GN=P2RY12	0.08035
GN=LMO7	0.00615	GN=RPL27A	0.03928	GN=KHDRBS1	0.08213
GN=LMO4	0.00698	GN=TNRC6C	0.04198	GN=SRSF3	0.08437
GN=SSBP3	0.00986	GN=RPS27A	0.04354	GN=DDX41	0.09079
GN=SYCP2L	0.01110	GN=LRRC16A	0.04709	GN=RPS13	0.09422

N.B. Only gene names were given for the putative UNR-interacting proteins to save table space due to the large number of suggested hits. This means that some identifiers are repeated due to different isoforms being suggested. The t-test p-values are for a two tailed paired t-test without multiple testing correction and increase from top to bottom for each column, starting on the left.

Table 4.21: Top ten putative SaOS-2 plus arsenite UNR interacting proteins by t-test p-value (using Progenesis data)

Protein	SaOS7	SaOS8	SaOS9	SaOS10	SaOS11	SaOS12	p-value
PFKL	1083	508	363	74500	71001	74644	0.00025
SETMAR	3937878	4202472	3573164	5430246	5815543	5147986	0.00052
IKBKAP	2085	2843	5216	81333	85611	95102	0.00139
ALK	234128	237994	236891	353285	358164	371992	0.00170
WDR1	15347	16102	32227	225332	263873	271295	0.00241
PCBP1	10192	10964	7337	922498	1101934	989078	0.00272
CCDC50	8577	7835	11619	515228	428615	503208	0.00312
HSPA6	274	367	1752	110832	131579	134049	0.00320
RNH1	199137	234258	218004	3683184	4546813	4031052	0.00385
TRMT1L	14279737	11323838	11379714	18564392	15513882	16510141	0.00432

N.B. The t-test p-values are for a two tailed paired t-test without multiple testing correction.

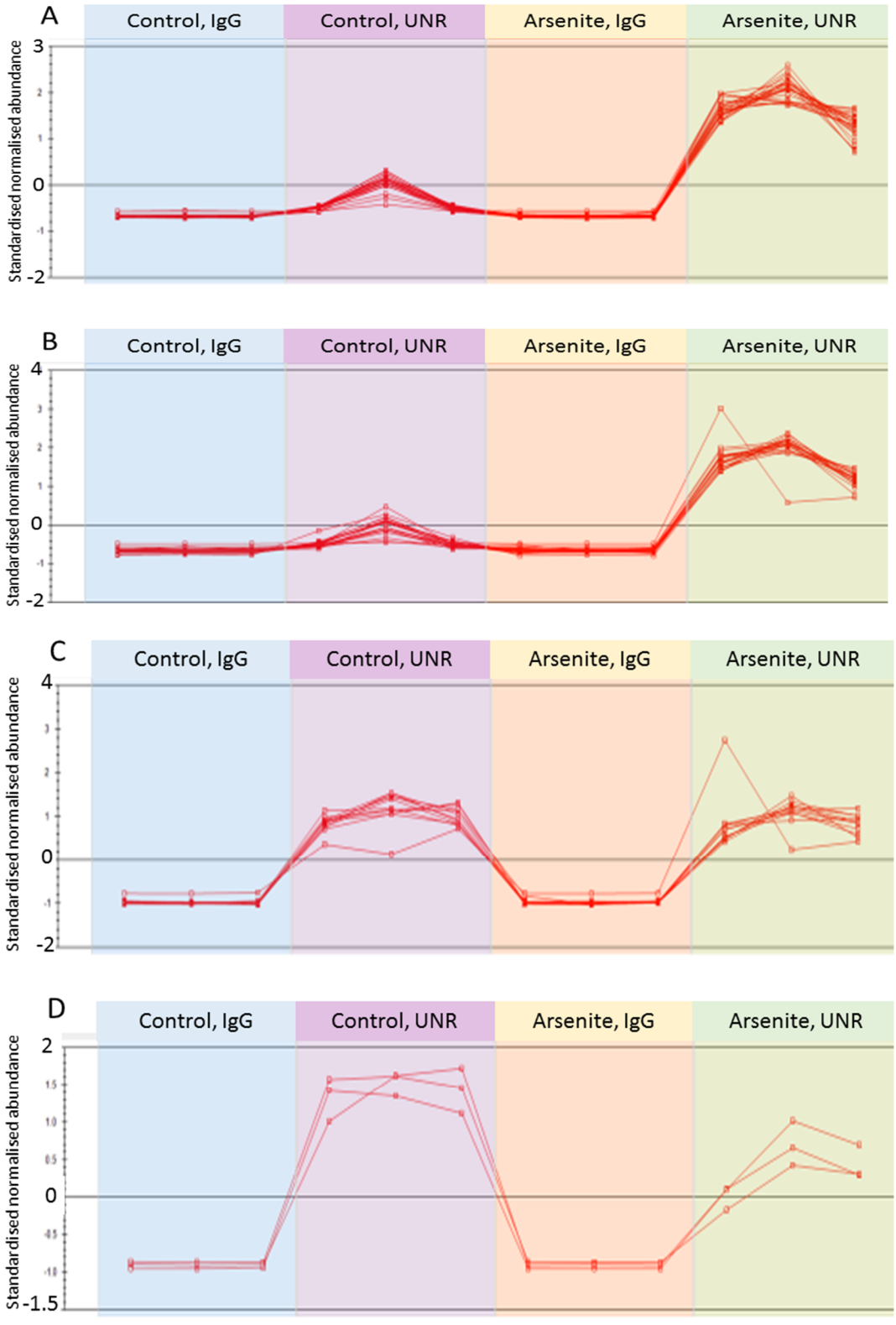


Figure 4.18: Standardised normalised abundances for peptides assigned to specific proteins (A=UNR ['different' forms combined], B = UNRIP, C = LDB1, D = LMO4) from the SaOS-2 data. The repeats in the blue column are IgG (minus arsenite), the repeats in the purple column are UNR (minus arsenite), the repeats in the orange column are IgG (plus arsenite) and the repeats in the green column are UNR (plus arsenite).

On this occasion, as well as UNR and UNRIP, it was decided to show peptide data for the LDB1 and LMO4. The peptides for each protein generally followed the same detection level pattern as each other with only one or two peptides deviating from the general pattern for a given protein (Figure 4.18). This implies that most of the peptides were correctly assigned to their given protein.

Finally, a fold change minimum of 10 was set as a significance cut off point (again only considering the data with ANOVA p-values of 0.05 or less). This gave a set of 158 putative UNR interactors for the minus arsenite samples following the removal of uncharacterised proteins (Table S2). The twelve SaOS-2 minus arsenite hits with infinite ratios, where all UNR pulldown values were non-zero, were tabulated in Table 4.22. There were 211 putative UNR interacting proteins suggested in SaOS-2 plus arsenite (Table S3). Of these, 24 had infinite ratios where all UNR pulldown values were >0. These are presented in Table 4.23.

Table 4.22: Putative SaOS-2 minus arsenite UNR interacting proteins with infinite ratios (using Progenesis data, all IgG=0, all UNR>0)

Protein	SaOS1	SaOS2	SaOS3	SaOS4	SaOS5	SaOS6	Ratio
TRA2B	0	0	0	4435	66929	109404	∞
DDX17	0	0	0	20405	43751	82584	∞
TUBA4A	0	0	0	5975	2818	61712	∞
NUMA1	0	0	0	10252	3348	29572	∞
TNRC6C	0	0	0	9070	19130	13192	∞
BMP15	0	0	0	1668	623	35507	∞
P4HB	0	0	0	2762	3967	23743	∞
RPS3A	0	0	0	1526	7190	17108	∞
SAP18	0	0	0	164	5828	19799	∞
LDB2	0	0	0	3602	5300	4232	∞
RPL8	0	0	0	217	6172	2664	∞
ANP32E	0	0	0	6	1626	709	∞

N.B. Ratios are mean(UNR)/mean(IgG) and all infinities were ignored unless every UNR value was greater than zero. The yellow shading highlights a protein absent in every IgG sample but present in every UNR sample.

Table 4.23: Putative SaOS-2 plus arsenite UNR interacting proteins with infinite ratios (using Progenesis data, all IgG=0, all UNR>0)

Description	SaOS7	SaOS8	SaOS9	SaOS10	SaOS11	SaOS12	ratio
PFN2	0	0	0	354772	473127	509770	∞
TBR1	0	0	0	127885	135678	343814	∞
NUMA1	0	0	0	148541	226850	205414	∞
LARS	0	0	0	12124	17343	268166	∞
DDX17	0	0	0	71030	69665	121418	∞
USMG5	0	0	0	10089	21651	184153	∞
ANP32E	0	0	0	34789	54185	115721	∞
FBXO45	0	0	0	39403	54938	22251	∞
NUP107	0	0	0	39403	54938	22251	∞
TTLL12	0	0	0	38540	24738	48882	∞
TBL1XR1	0	0	0	34212	38801	23504	∞
PPIA	0	0	0	12439	5854	77771	∞
FNDC3B	0	0	0	24194	34413	22260	∞
TNRC6C	0	0	0	29062	25800	19529	∞
MAGED1	0	0	0	19160	42866	10798	∞
SUMO1	0	0	0	7477	46151	16560	∞
RPS3A	0	0	0	2396	35868	23634	∞
ABCE1	0	0	0	5212	19789	30993	∞
ARF4	0	0	0	10906	3557	39720	∞
SRSF2	0	0	0	2308	31211	2710	∞
GNB2L1	0	0	0	3319	8353	22353	∞
BMP15	0	0	0	4030	5669	19563	∞
RPS9	0	0	0	1882	20677	4400	∞
LDB2	0	0	0	1336	4452	3650	∞

N.B. Ratios are mean(UNR)/mean(IgG) and all infinities were ignored unless every UNR value was greater than zero. The yellow shading highlights a protein absent in every IgG sample but present in every UNR sample.

4.11.1 Discussion of SaOS-2 RIP dataset comparison to the HeLa and U2OS datasets

The top ten putative UNR interacting proteins by t-test p-value for SaOS-2 minus arsenite were all different to those for unstressed U2OS (Tables 4.14, 4.20). There was some cross-over between the top ten hits for unstressed SaOS-2 and unstressed HeLa (Tables 4.7A, 4.20). LDB1, SSBP3 and C4A were top ten hits for both SaOS-2 and HeLa (Tables 4.7A, 4.20).

As with the top ten putative UNR interacting proteins in unstressed SaOS-2 and U2OS, there was no cross-over between the top ten putative UNR interactors by t-test p-value for arsenite treated SaOS-2 and stressed U2OS (Tables 4.15, 4.21). The cross-over between the top ten hits for stressed SaOS-2 and stressed HeLa included ALK and RNH1 (Tables 4.8A, 4.21).

Whereas the top ten hits by UNR/IgG ratio were presented for HeLa and U2OS, there were more than ten infinite ratios for both stressed and unstressed SaOS-2. All infinite ratios where all three UNR samples had non-zero Progenesis quantification values were presented. Having taken that caveat into consideration, comparison was then made between the infinite ratio hits for SaOS-2 and the top ten hits for the same arsenite treatment state in the other two cell types.

For the unstressed samples, there was no cross-over between top hits for U2OS and SaOS-2 (Tables 4.16, 4.22). Likewise, there was no cross-over between top hits for HeLa and SaOS-2 (Tables 4.9A, 4.22). For the stressed samples, PPIA was an infinite ratio hit for both U2OS and SaOS-2 although there was no other cross-over between the top hits lists for the two cell types (Tables 4.17, 4.23). There was no cross-over between the top lists for stressed HeLa and SaOS-2 (Tables 4.10A, 4.23).

The lack of appreciable cross-over is interesting but not unexpected. In order to get a better idea of hits that were shared between cell types, all putative UNR-interacting proteins from the minus arsenite samples with a unique gene ID and t-test p-value <0.1 were aligned by cell type and Microsoft Excel was used to highlight duplicates. Of 21 putative hits in HeLa, 15 were present in one of the other two cell types and five were present in both the other cell types (C4A, LDB1,

NARR, SQSTM1 and SRSF3). Of 36 putative UNR interactors in SaOS-2, 14 were present in at least one of the other cell types. Of 61 U2OS hits, 16 were present in at least one of the other cell types.

That analysis was then repeated with the plus arsenite data. Of 58 putative UNR interactors in stressed HeLa cells, 41 were present in at least one of the osteosarcoma cell lines and ten were present in all three cell types (C4A, EWSR1, HNRNPK, HUWE1, LDB1, NARR, PCBP2, RPS2, TRIP6 and UNR itself). Of 42 hits in U2OS, 19 were present in at least one other cell type and, of 142 hits in SaOS-2, 42 were present in at least one other cell type. C4A, LDB1 and NARR were hits in all cell types, both stressed and unstressed. It was noted that HeLa hits were more reproducible than hits in other the other cell types.

4.11.2 Progenesis results for SaOS-2 RIP dataset exported to AmiGO 2

The plus and minus arsenite hits suggested by Progenesis were exported separately to AmiGO 2 version 2.3.2 following removal of NARR and heavy and light chains (for both plus and minus arsenite) that were assigned to non-IgG antibodies. Other than that, processing was as per the HeLa samples (see section 4.9.4).

In each case, AmiGO 2 failed to recognise three proteins (suggested functions from Uniprot):

H3BN98 (a possible structural component of the ribosome that interacts with MYH9),

B4DLN1 (a possible structural component of the ribosome that interacts with subunits of the mitochondrial ribosome),

F5H423 (that may be involved in binding to GTP and small GTPase mediated signalling activity).

4.11.3 GO term results for SaOS-2 minus arsenite

A total of 171 individual proteins were entered into AmiGO 2 but three were not recognised by the tool and were removed from further consideration.

A Bonferroni multiple testing corrected p-value of 0.05 was chosen as the significance cut off point and the top ten hits by adjusted p-value were recorded in Table 4.24 (Table 4.24A = biological process, Table 4.24B = molecular function, Table 4.24C = cellular component).

Table 4.24A: Top ten enriched biological process GO terms, by Bonferroni adjusted p-value (SaOS-2 minus arsenite samples)

GO biological process complete	Ref	Obs	Exp	FE	p-value
mRNA metabolic process (GO:0016071)	625	40	4.98	8.02	1.15E-20
RNA processing (GO:0006396)	930	45	7.42	6.07	6.24E-19
single-organism intracellular transport (GO:1902582)	1372	53	10.94	4.84	1.05E-18
intracellular transport (GO:0046907)	1529	55	12.19	4.51	3.55E-18
establishment of localization in cell (GO:0051649)	1670	56	13.32	4.2	3.52E-17
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	111	20	0.89	22.59	4.28E-17
selenium compound metabolic process (GO:0001887)	114	20	0.91	22	7.15E-17
cotranslational protein targeting to membrane (GO:0006613)	114	20	0.91	22	7.15E-17
protein targeting to ER (GO:0045047)	114	20	0.91	22	7.15E-17
protein localization to endoplasmic reticulum (GO:0070972)	138	21	1.1	19.08	1.29E-16

Table 4.24B: Top ten enriched molecular function GO terms, by Bonferroni adjusted p-value (SaOS-2 minus arsenite samples)

GO molecular function complete	Ref	Obs	Exp	FE	p-value
poly(A) RNA binding (GO:0044822)	1165	67	9.29	7.21	3.41E-36
RNA binding (GO:0003723)	1718	77	13.7	5.62	2.03E-35
heterocyclic compound binding (GO:1901363)	6512	120	51.94	2.31	7.09E-24
organic cyclic compound binding (GO:0097159)	6581	120	52.49	2.29	2.02E-23
protein binding (GO:0005515)	10761	149	85.82	1.74	1.78E-22
binding (GO:0005488)	14353	163	114.47	1.42	3.10E-19
nucleic acid binding (GO:0003676)	4514	93	36	2.58	1.97E-18
small molecule binding (GO:0036094)	3096	71	24.69	2.88	1.43E-14
nucleotide binding (GO:0000166)	2772	67	22.11	3.03	1.44E-14
nucleoside phosphate binding (GO:1901265)	2773	67	22.12	3.03	1.46E-14

Table 4.24C: Top ten enriched cellular component GO terms, by Bonferroni adjusted p-value (SaOS-2 minus arsenite samples)

GO cellular component complete	Ref	Obs	Exp	FE	p-value
intracellular organelle part (GO:0044446)	7968	132	63.55	2.08	2.82E-24
macromolecular complex (GO:0032991)	4651	102	37.09	2.75	4.49E-24
extracellular exosome (GO:0070062)	2725	78	21.73	3.59	3.86E-23
organelle part (GO:0044422)	8176	132	65.21	2.02	4.90E-23
extracellular vesicle (GO:1903561)	2738	78	21.84	3.57	5.26E-23
extracellular organelle (GO:0043230)	2740	78	21.85	3.57	5.51E-23
ribonucleoprotein complex (GO:1990904)	750	42	5.98	7.02	1.56E-20
intracellular ribonucleoprotein complex (GO:0030529)	750	42	5.98	7.02	1.56E-20
organelle (GO:0043226)	12818	157	102.23	1.54	4.83E-20
vesicle (GO:0031982)	3642	84	29.05	2.89	3.37E-19

4.11.4 Consideration of GO term results for SaOS-2 minus arsenite

More than half of the ten most significant biological process GO terms, using putative UNR-interacting proteins from unstressed SaOS-2 cells, were related to intracellular transport and localisation (Table 4.24A). The top two hits by p-value were related to mRNA and RNA processing. As seen with the U2OS minus arsenite GO term analysis, selenium compound metabolic processes was also a top biological process GO term with SaOS-2 cells (Tables 4.18A, 4.24A). Many of the top ten molecular function and cellular component GO terms were related to RNA or very general terms like protein binding, binding and organelle, all of which have over 10000 proteins annotated to them in the reference set. Interestingly, 78 putative UNR-interacting proteins were annotated to extracellular exosome, extracellular vesicle and extracellular organelle (presumably exactly the same proteins in each case, although that wasn't specifically confirmed). This is interesting on several levels. Firstly, previous work in the lab had provided hints that UNR may be present in vesicles that had blebbed off cells undergoing apoptosis. On the other hand, it is not immediately obvious why UNR would be interacting with extracellular proteins outside the cell in unstressed cells. One consideration is that cancer cells have been reported to usurp exosome function to change the proteome or expel intracellular drugs, as reviewed in (Azmi *et al.*, 2013). It is possible that UNR was being expelled from the cell

or being hijacked to cause the expulsion of other proteins or RNAs. This could potentially become an interesting new avenue in UNR-related research.

4.11.5 GO term results for SaOS-2 plus arsenite

A total of 219 individual proteins were entered into AmiGO 2 but three were not recognised by the tool and were removed from further consideration.

A Bonferroni multiple testing corrected p-value of 0.05 was chosen as the significance cut off point and the top ten hits by adjusted p-value were recorded in Table 4.25 (Table 4.25A = biological process, Table 4.25B = molecular function, Table 4.25C = cellular component).

Table 4.25A: Top ten enriched biological process GO terms, by Bonferroni adjusted p-value (SaOS-2 plus arsenite samples)

GO biological process complete	Ref	Obs	Exp	FE	p-value
mRNA metabolic process (GO:0016071)	625	50	6.49	7.71	1.60E-25
selenium compound metabolic process (GO:0001887)	114	28	1.18	23.67	1.78E-25
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	111	27	1.15	23.44	2.39E-24
cotranslational protein targeting to membrane (GO:0006613)	114	27	1.18	22.82	4.78E-24
protein targeting to ER (GO:0045047)	114	27	1.18	22.82	4.78E-24
RNA processing (GO:0006396)	930	57	9.65	5.91	6.32E-24
establishment of protein localization to endoplasmic reticulum (GO:0072599)	118	27	1.22	22.05	1.17E-23
single-organism intracellular transport (GO:1902582)	1372	67	14.24	4.71	2.03E-23
peptide biosynthetic process (GO:0043043)	539	45	5.59	8.04	2.74E-23
protein localization to endoplasmic reticulum (GO:0070972)	138	28	1.43	19.55	3.03E-23

Table 4.25B: Top ten enriched molecular function GO terms, by Bonferroni adjusted p-value (SaOS-2 plus arsenite samples)

GO molecular function complete	Ref	Obs	Exp	FE	p-value
RNA binding (GO:0003723)	1718	107	17.83	6	1.71E-53
poly(A) RNA binding (GO:0044822)	1165	91	12.09	7.53	1.01E-51
heterocyclic compound binding (GO:1901363)	6512	163	67.58	2.41	4.75E-37
organic cyclic compound binding (GO:0097159)	6581	163	68.3	2.39	2.05E-36
nucleic acid binding (GO:0003676)	4514	130	46.84	2.78	7.98E-31
protein binding (GO:0005515)	10761	193	111.67	1.73	4.78E-29
binding (GO:0005488)	14353	211	148.95	1.42	2.70E-24
nucleotide binding (GO:0000166)	2772	86	28.77	2.99	9.64E-19
nucleoside phosphate binding (GO:1901265)	2773	86	28.78	2.99	9.87E-19
structural molecule activity (GO:0005198)	775	45	8.04	5.6	1.60E-17

Table 4.25C: Top ten enriched cellular component GO terms, by Bonferroni adjusted p-value (SaOS-2 plus arsenite samples)

GO cellular component complete	Ref	Obs	Exp	FE	p-value
extracellular exosome (GO:0070062)	2725	108	28.28	3.82	9.84E-36
extracellular vesicle (GO:1903561)	2738	108	28.41	3.8	1.52E-35
extracellular organelle (GO:0043230)	2740	108	28.43	3.8	1.63E-35
macromolecular complex (GO:0032991)	4651	131	48.27	2.71	1.70E-30
vesicle (GO:0031982)	3642	115	37.8	3.04	1.84E-29
ribonucleoprotein complex (GO:1990904)	750	56	7.78	7.19	1.82E-28
intracellular ribonucleoprotein complex (GO:0030529)	750	56	7.78	7.19	1.82E-28
intracellular organelle part (GO:0044446)	7968	164	82.69	1.98	3.12E-26
extracellular region part (GO:0044421)	3745	111	38.86	2.86	1.59E-25
organelle part (GO:0044422)	8176	165	84.85	1.94	1.87E-25

4.11.6 Consideration of GO term results for SaOS-2 plus arsenite

The SaOS-2 plus arsenite samples generated more putative UNR interactors than any of the other conditions. The top two most enriched biological process GO terms were mRNA metabolic processes and selenium compound metabolic process. Whilst GO terms related to RNA were to be expected, it is worth considering some of the numbers associated with the selenium-related GO term. There were 114 proteins annotated to the selenium compound metabolic process GO term. Of these, 5 were putative UNR interactors detected in unstressed HeLa cells (Table 4.12A). Ten suggested UNR interacting proteins from stressed HeLa cells were annotated to the GO term (Table 4.13A). Eleven of the annotated proteins were flagged as putative UNR-interactors in U2OS minus arsenite cells (Table 4.18A). Whilst it didn't make the top ten biological process GO terms by p-value from the U2OS plus arsenite cells, 15 putative UNR interactors were annotated to the GO term (data not shown). Twenty were flagged as possible UNR interactors in SaOS-2 minus arsenite. The latest condition considered (SaOS-2 plus arsenite) had 28 of the 114 proteins annotated to selenium compound metabolic process as putative UNR-interactors (Table 4.25A). The Bonferroni-adjusted p-value associated with the GO term for SaOS-2 plus arsenite was 1.78×10^{-25}. The differences in the number of putative UNR interactors annotated to the GO term in different cell types/stress conditions retrospectively supports having used different cell types in this experiment. It seems highly likely that UNR regulates selenium compound metabolism in

some way, but that finding may have been overlooked if only HeLa cells had been used. Other top biological process GO terms for SaOS-2 plus arsenite include terms relating to translation and localisation (Table 4.25A).

The top molecular function GO terms associated with potential UNR-interacting proteins from SaOS-2 plus arsenite are largely related to RNA/nucleotide binding (Table 4.25B).

Whereas the SaOS-2 minus arsenite samples yielded 78 putative UNR-interacting proteins annotated to extracellular exosome, SaOS-2 plus arsenite yielded 108, making it the top cellular component GO term by p-value, 9.84×10^{-36}, (Table 4.25C). Another top cellular component GO term was the more expected, based on the literature, ribonucleoprotein complex (Table 4.25C).

4.12 Validation of selected putative UNR interacting proteins

4.12.1 HUWE1

In order to validate one of the putative UNR interacting proteins that had been consistently flagged up across each of the cell types examined, HUWE1, an IP-Western was carried out to ascertain if it could be pulled down with UNR from HeLa cell lysate.

The IP was carried out using the RIP protocol described previously using lysates left over from the main experiments. Two sets of two RIPs were undertaken, one set using lysate from unstressed HeLa cells and the other using arsenite treated HeLa cell lysate. Each set contained two IPs – one with an antibody against UNR and the other with a conspecific control IgG. 5% of the total lysate used was retained from each set to run an input lane on a Western. The remainder of the lysate was divided equally between the UNR and control IgG IPs. The resultant Western was quite dirty in the HUWE1 region (Figure 4.19).

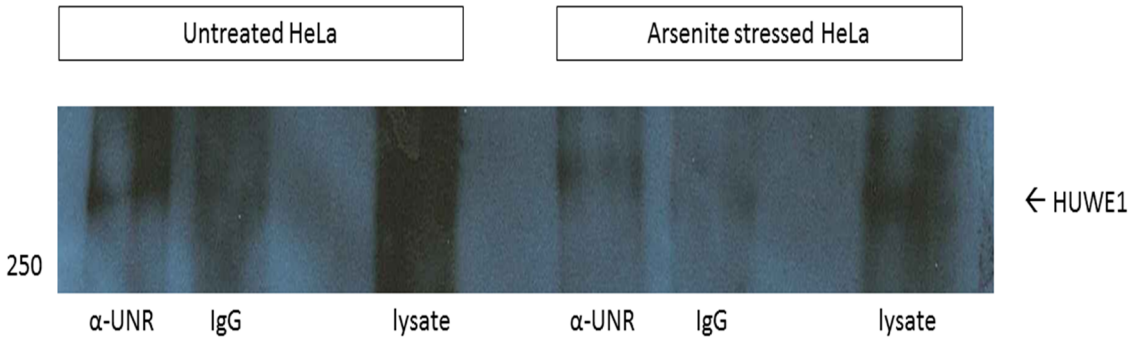


Figure 4.19: Result of an IP-Western probing for HUWE1 in which an anti-UNR immunoglobulin or a conspecific control IgG, as stated, was used to pulldown proteins from cell lysates that had been treated with 1 mM of sodium arsenite, or a similar volume of sterile PBS, in fresh DMEM for 1 hour immediately prior to harvesting (as stated). Input lanes containing 5% of the total amount of lysate used in the experiment were loaded in each case (labelled "lysate"). The immunoprecipitated proteins were boiled in loading buffer and run on a 10% polyacrylamide gel prior to being transferred onto nitrocellulose membrane which was blocked, probed with an anti-HUWE1 immunoglobulin raised in rabbit (see section 2.1.1) followed by a HRP-conjugated anti-rabbit IgG secondary antibody and then developed using an enhanced chemiluminescence-based approach. The number on the left hand side of the image shows approximate molecular masses (in kilodaltons), as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher). This experiment was only carried out on one occasion due to time constraints at the end of the project.

Whilst it looked like there may be differential binding to UNR over IgG, it was decided to look for co-localisation of UNR and HUWE1 by immunofluorescence microscopy across all three cell types in search of further corroboration (Figures 4.20-4.26).

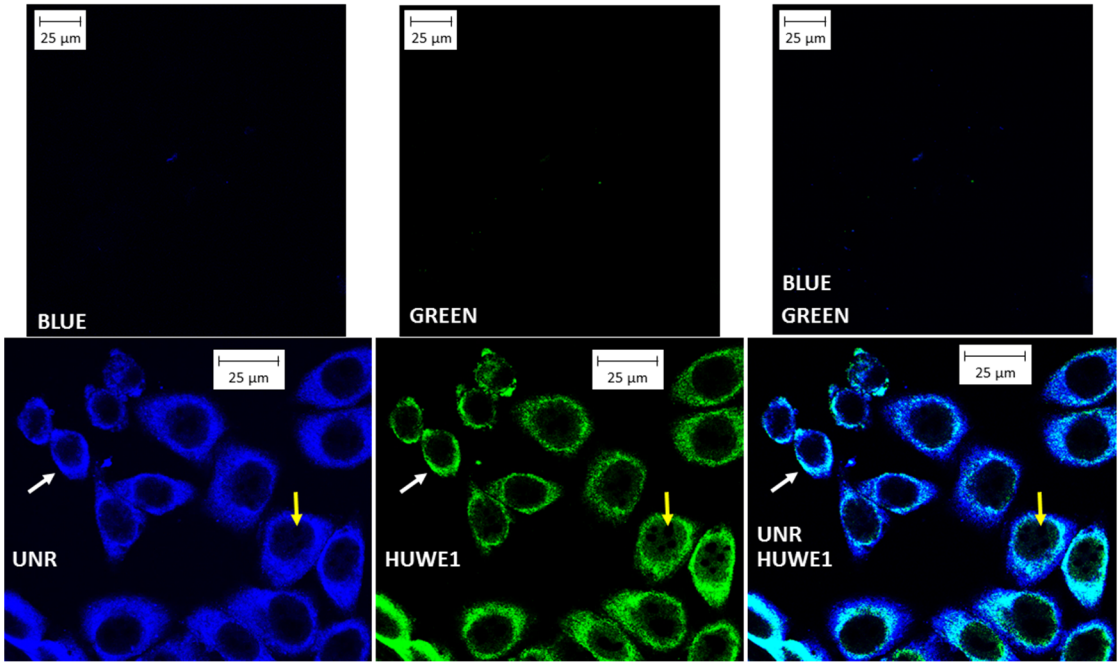


Figure 4.20: Confocal immunofluorescence microscopy images of non-arsenite treated HeLa cells fixed and stained immediately after having been mock treated with sterile PBS in fresh DMEM for 1 hour. The upper images show background staining from the fluorophore-conjugated secondary antibodies, as coloured, in the absence of primary antibodies. The lower images show staining for UNR (blue), HUWE1 (green), and an overlaid image showing staining for both UNR and HUWE1. The white arrows show an area of apparent diffuse co-localisation between UNR and HUWE1. The yellow arrows demonstrate that neither UNR nor HUWE1 were visible in structures that had the appearance of nucleoli, although the lack of staining for a nucleolar marker makes it impossible to state this with certainty. The scale bars are all 25 µm.

Diffuse co-localisation of UNR and HUWE1 was observed in unstressed HeLa cells. Both proteins were predominantly cytoplasmic in distribution although some staining for both was visible in the nucleus outside of structures that appeared to be nucleoli (Figure 4.20).

Punctate co-localisation of UNR and HUWE1 was observed in arsenite stressed HeLa cells (Figure 4.21). These regions were believed to be stress granules.

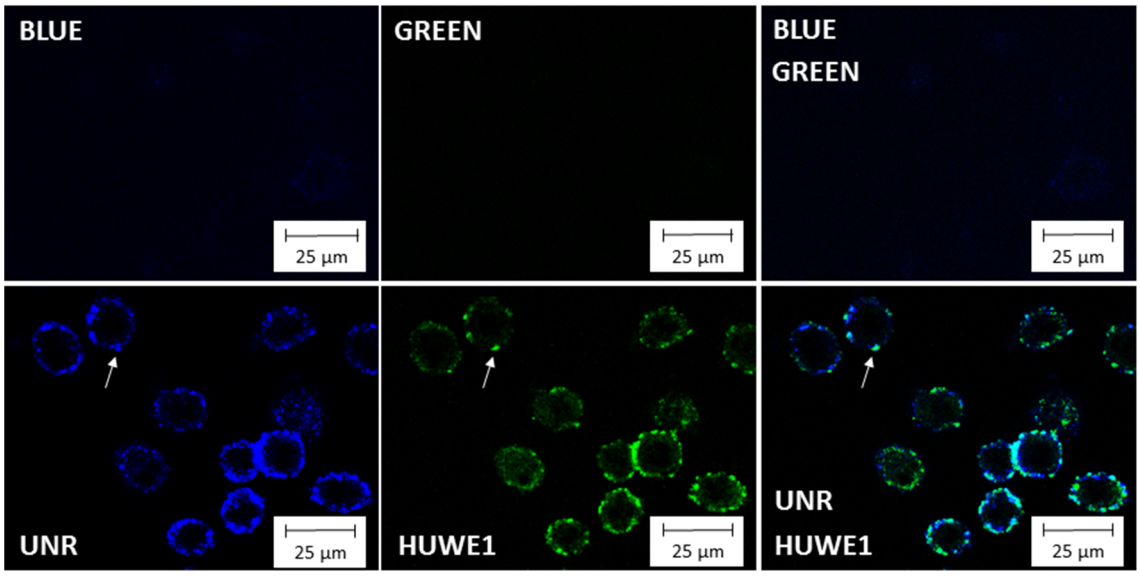


Figure 4.21: Confocal immunofluorescence microscopy images of arsenite treated HeLa cells fixed and stained immediately after having been treated with 1 mM sodium arsenite in fresh DMEM for 1 hour. The upper images show background staining from the fluorophore-conjugated secondary antibodies, as coloured, in the absence of primary antibodies. The lower images show staining for UNR (blue), HUWE1 (green), and an overlaid image showing staining for both UNR and HUWE1. The white arrows show a punctate area of co-localisation between UNR and HUWE1. The scale bars are all 25 µm.

Whilst co-localisation by immunofluorescence does not prove that there is a direct, or even indirect, interaction between UNR and HUWE1, it does support the mass spectrometry data that suggested such a link.

There was an interesting co-localisation pattern between UNR and HUWE1 in unstressed U2OS cells in which there was a large amount of diffuse co-localisation that appeared particularly pronounced at some cell-cell junctions (Figure 4.22).

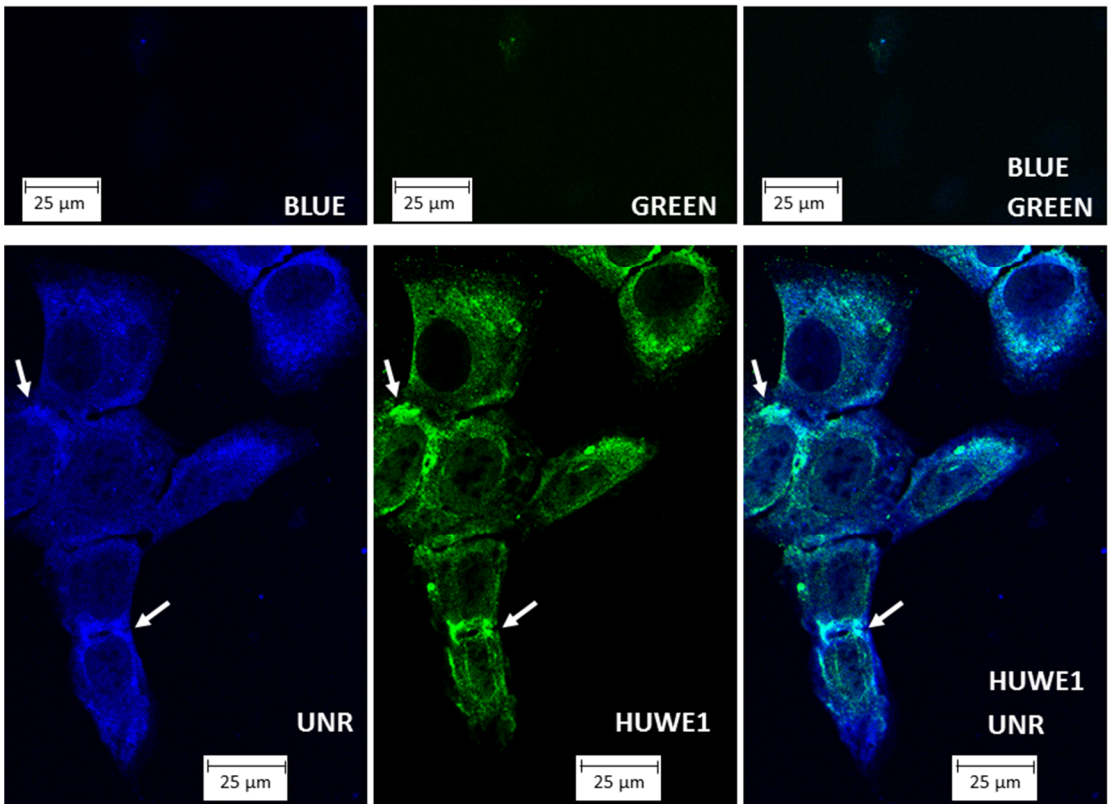


Figure 4.22: Confocal immunofluorescence microscopy images of non-arsenite treated U2OS cells fixed and stained immediately after having been mock treated with sterile PBS in fresh DMEM for 1 hour. The upper images show background staining from the fluorophore-conjugated secondary antibodies, as coloured, in the absence of primary antibodies. The lower images show staining for UNR (blue), HUWE1 (green), and an overlaid image showing staining for both UNR and HUWE1. The white arrows show areas of apparent diffuse co-localisation between UNR and HUWE1 at cell extremities or areas of physical cell-cell interaction. The scale bars in all images are 25 µm.

Diffuse co-localisation of UNR and HUWE1 at cell-cell junctions had been noted in arsenite-stressed HeLa cells that appeared to be undergoing cytokinesis (Figure 4.23).

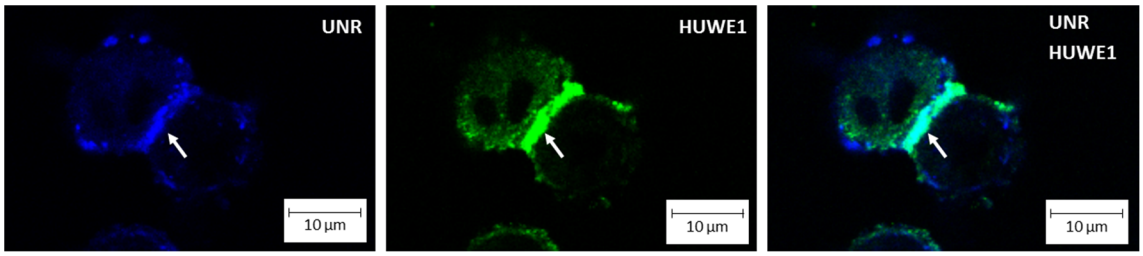


Figure 4.23: Confocal immunofluorescence microscopy images of arsenite treated HeLa cells fixed and stained immediately after having been treated with 1 mM sodium arsenite in fresh DMEM for 1 hour. The images show staining for UNR (blue), HUWE1 (green), and an overlaid image showing staining for both UNR and HUWE1. The white arrows show an area of diffuse co-localisation between UNR and HUWE1, possibly in a cytokinetic contractile ring. The scale bars are 10 µm.

It was noted previously that adherens junction and anchoring junction were on the top ten list of enriched cellular component GO terms for U2OS, albeit from the plus arsenite samples (Table 4.19D). Taken together, these observations support the idea that UNR and HUWE1 co-localise at cell-cell junctions under certain conditions and are possibly present in the contractile ring during cytokinesis.

As in arsenite-treated HeLa cells (Figure 4.21), UNR and HUWE1 appear to co-localise to stress granules in arsenite treated U2OS cells (Figure 4.24). Unlike unstressed U2OS cells (Figure 4.22), however, there did not appear to be areas of co-localisation at cell-cell junctions in arsenite-stressed U2OS cells (Figure 4.24). It was noted previously in the lab that arsenite treatment appears to cause the breakdown of cell-cell interactions (data not shown).

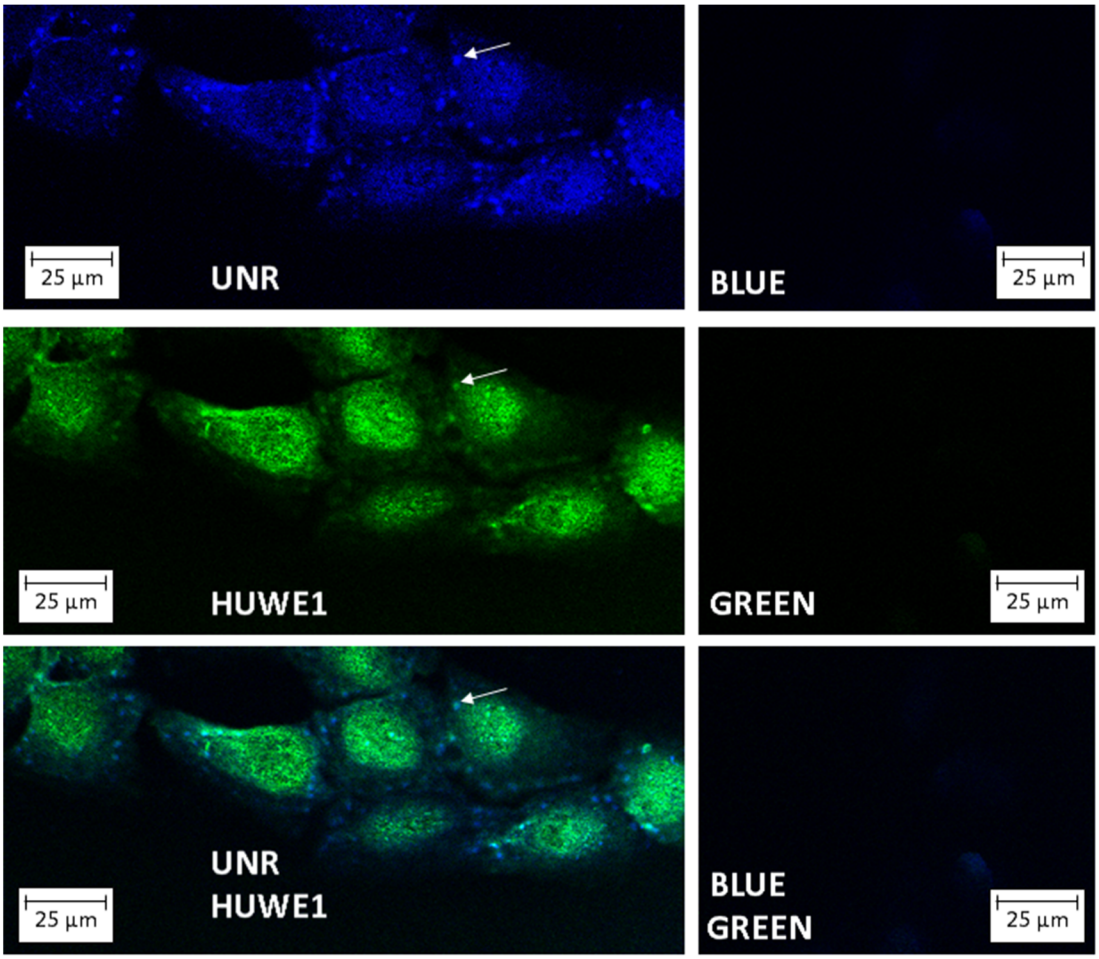


Figure 4.24: Confocal immunofluorescence microscopy images of arsenite treated U2OS cells fixed and stained immediately after having been treated with 1 mM sodium arsenite in fresh DMEM for 1 hour. The right hand images show background staining from the fluorophore-conjugated secondary antibodies, as coloured, in the absence of primary antibodies. The left hand images show staining for UNR (blue), HUWE1 (green), and an overlaid image showing staining for both UNR and HUWE1. The white arrows show an area of punctate co-localisation of UNR and HUWE1. The scale bar in all images is 25 µm.

Some areas of apparent co-localisation between UNR and HUWE1 were observed in unstressed SaOS-2 cells. There were both punctate areas of co-localisation (Figure 4.25, white arrows) and larger areas of co-localisation around the nucleus (Figure 4.25, yellow arrows). UNR had previously been recorded to be concentrated around the nucleus and endoplasmic reticulum by immunofluorescence microscopy in the MCF7 cell line (Jacquemin-Sablon *et al.*, 1994). Whilst this is interesting in terms of seeing UNR concentrated around the nucleus, in general, the distribution of UNR in the cell types used during this project has been quite different to the images obtained by Jacquemin-Sablon.

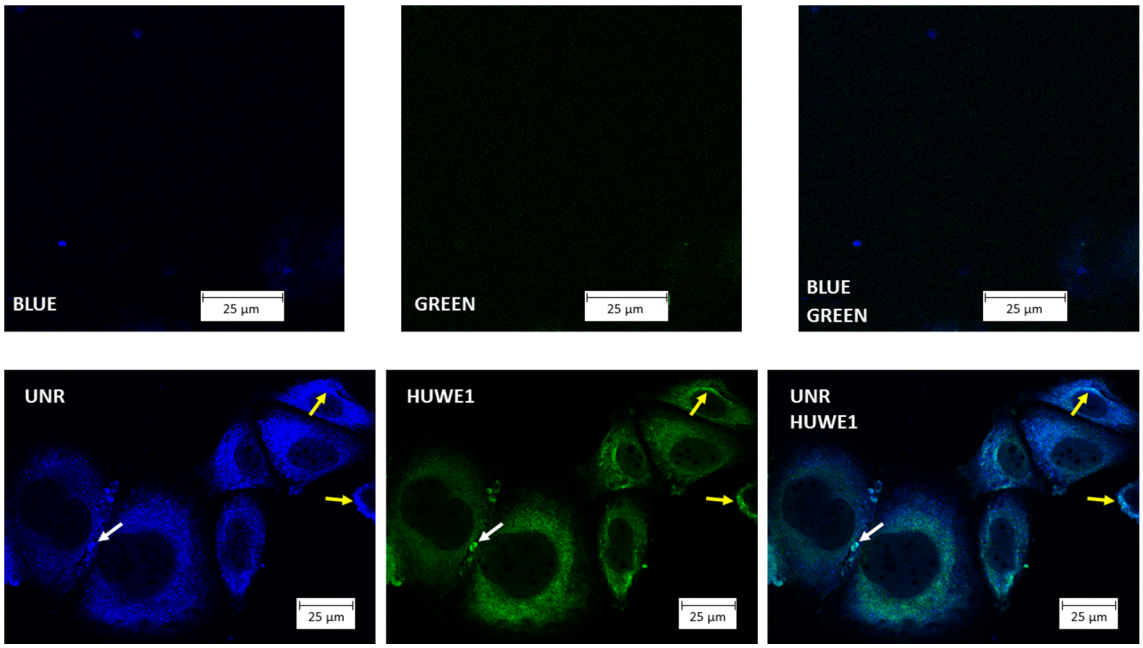


Figure 4.25: Confocal immunofluorescence microscopy images of non-arsenite treated SaOS-2 cells fixed and stained immediately after having been mock treated with sterile PBS in fresh DMEM for 1 hour. The upper images show background staining from the fluorophore-conjugated secondary antibodies, as coloured, in the absence of primary antibodies. The lower images show staining for UNR (blue), HUWE1 (green), and an overlaid image showing staining for both UNR and HUWE1. The white arrows show punctate areas of co-localisation between UNR and HUWE1. The yellow arrows show larger areas of UNR co-localisation with HUWE1 near the nucleus. The scale bar in all images is 25 µm.

Arsenite-stressed SaOS-2 cells also showed some areas of apparent co-localisation of UNR and HUWE1. The co-localisation appeared to be diffuse even though UNR had adopted a more punctate localisation (Figure 4.26, yellow arrows). Interestingly, there were clear stress granule-like structures that contained UNR but did not contain HUWE1 (Figure 4.26, white arrows).

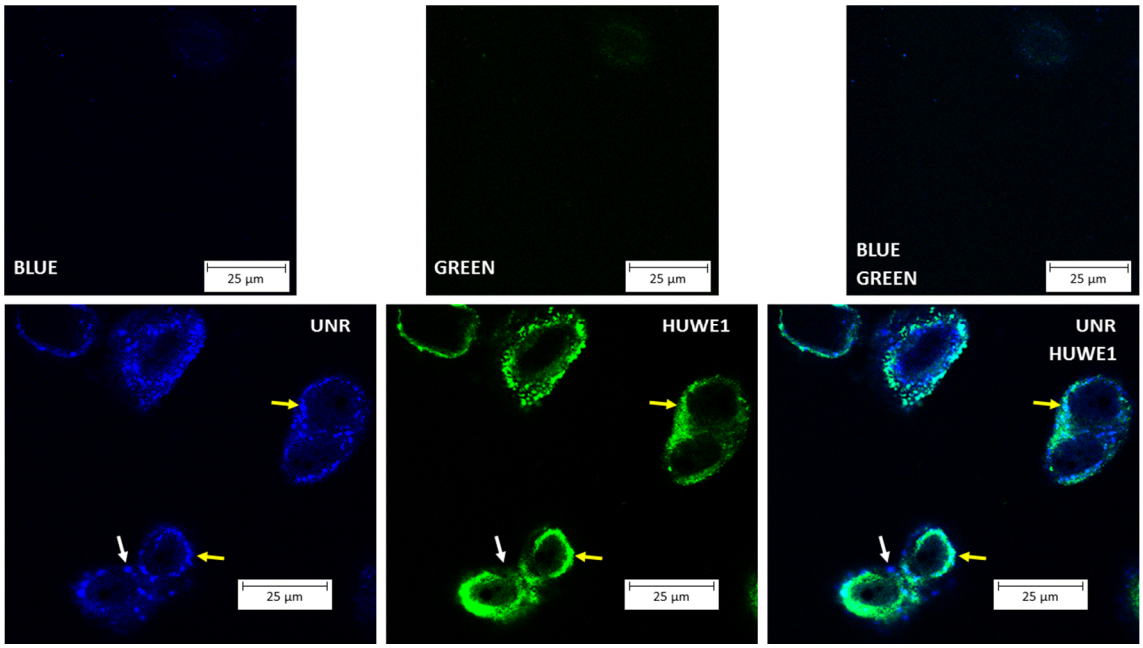


Figure 4.26: Confocal immunofluorescence microscopy images of arsenite treated SaOS-2 cells fixed and stained immediately after having been treated with 1 mM sodium arsenite in fresh DMEM for 1 hour. The upper images show background staining from the fluorophore-conjugated secondary antibodies, as coloured, in the absence of primary antibodies. The lower images show staining for UNR (blue), HUWE1 (green), and an overlaid image showing staining for both UNR and HUWE1. The yellow arrows show areas of co-localisation of UNR and HUWE1. The white arrows show a punctate area of UNR staining without any HUWE1. The scale bar in all images is 25 µm.

It is interesting to consider the TP53 status of the cell lines used here. Whilst at least one subclone of HeLa (HeLa S3) lacks the TP53 gene, other HeLa cells do possess it (Jia *et al.*, 1997). The exact identity of the HeLa cells used here was unknown. It was nevertheless clear that they did express TP53 as the protein was observed, for example, by immunofluorescence microscopy (e.g. Figure 3.8). It should also be considered that the HeLa cell line was transformed with genetic information from human papilloma virus type 16. The viral E6 protein is expressed in HeLa and can both repress the transcriptional output of TP53 (Hoppe-Seyler & Butz, 1993) and target the protein itself for proteosomal degradation (Scheffner *et al.*, 1990). SaOS-2 has been shown to have lost the TP53 gene whereas the related U2OS cell line retains the gene and expresses the protein at equivalent levels to that seen in fibroblasts (Diller *et al.*, 1990). It should be noted that U2OS also overexpress HDM2 (Florenes *et al.*, 1994). Whilst TP53 expression in U2OS cells has been shown in this thesis by Western blotting (Figure 3.15), it is still possible that HDM2 could be acting to prevent TP53 transactivation or interaction with transcriptional co-

activators (discussed in Shi & Gu 2012). To summarise, SaOS-2 does not contain any TP53 (gene, transcript or protein) whereas both U2OS and HeLa express TP53 transcripts and protein. For this reason, differences in UNR distribution between SaOS-2 and both U2OS and, though to a lesser extent, HeLa could be related to TP53. On the other hand, beyond other differences between the cell type proteomes, any comparison of TP53 status is confounded with Hdm2 over-expression/E6 expression. Further research is required to ascertain which, if any, of the myriad cellular functions of TP53 could be involved.

4.12.2 Sequestosome-1

Having some evidence that UNR really does interact with HUWE1, it was decided to carry out an IP-Western on another protein that had been suggested as a top putative UNR-interactor across all three cell types, SQSTM1, using arsenite-stressed and unstressed HeLa cell lysates (Figure 4.27).

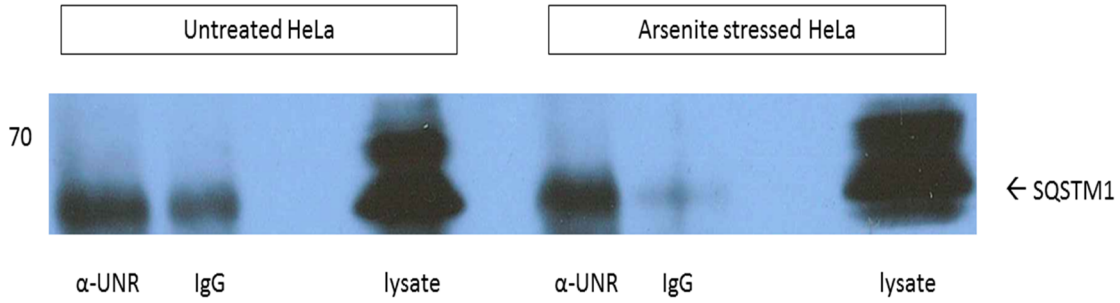


Figure 4.27: Result of an IP-Western in which an anti-UNR immunoglobulin or a conspecific control IgG, as stated, was used to pulldown proteins from cell lysates that had been treated with 1 mM of sodium arsenite, or a similar volume of sterile PBS, in fresh DMEM for 1 hour immediately prior to harvesting (as stated). Input lanes containing 5% of the total amount of lysate used in the experiment were loaded in each case (labelled lysate). The immunoprecipitated proteins were boiled in loading buffer and run on a 10% polyacrylamide gel prior to being transferred onto nitrocellulose membrane which was blocked, probed with an anti-SQSTM1 immunoglobulin raised in goat (see section 2.1.1) followed by a HRP-conjugated anti-goat IgG secondary antibody and then developed using an enhanced chemiluminescence-based approach. The number on the left hand side of the image is an approximate molecular mass (in kilodaltons), as estimated using PageRuler Plus Prestained Protein Ladder (ThermoFisher).

The IP-Western showed clear binding of SQSTM1 to UNR over the control IP, especially in arsenite treated HeLa cells, thereby validating SQSTM1 as a UNR-interacting protein (at least in HeLa cells). Whilst only a small selection of the total number of novel putative UNR-interacting proteins suggested by the work in this chapter, seeing evidence of a likely interaction in the two that were actually tested was promising. Whilst it remains to be seen whether the other suggested hits are genuine or not, evidence has been provided showing that not all the putative interactors are false positives.

4.13 Summary of chapter

Immunoprecipitations were carried out with an antibody against UNR in three different cell types (HeLa, U2OS and SaOS-2), both arsenite stressed and unstressed. The immunoprecipitate samples were then separated and a portion was subjected to analysis by mass spectrometry in an attempt to locate some previously unknown UNR interacting proteins.

Initial trial runs using unstressed HeLa cell lysates implied that there were some reproducible hits although it was not known how significant this was as it was not known how many proteins could be located by the mass spectrometer. Later, an estimate of over 2000 proteins was obtained using *UNR* knockdown data (number of proteins identified by Scaffold with a minimum of 2 peptides when the peptide cut off was 95% and the protein cut off was 99%, data not shown).

Paired t-tests were carried out on proteins quantified using the Progenesis software. As the t-test p-values were not strong enough, a multiple testing correction was not used. Validation was instead left to protein hit reproducibility between conditions and whether or not strong p-values were obtained for significantly overrepresented GO terms.

Among the most reproducible protein hits were UNRIP, which was expected, HUWE1, NARR, SQSTM1 and LDB1. HUWE1 was validated by IP-Western blot and immunofluorescence microscopy and SQSTM1 was validated by IP-Western blot.

The GO term overrepresentation data were extremely promising. The top biological process GO term for all conditions was ‘mRNA metabolic process’ (Table 4.26).

Table 4.26: Bonferroni-adjusted p-value for the overrepresentation of the 'mRNA metabolic process' biological process GO term across all conditions

Top biological process GO term	Cell type	Arsenite treatment	Bonferroni-adjusted p-value
mRNA metabolic process	HeLa	No	4.34×10^{-8}
mRNA metabolic process	HeLa	Yes	1.17×10^{-14}
mRNA metabolic process	U2OS	No	4.87×10^{-21}
mRNA metabolic process	U2OS	Yes	1.24×10^{-24}
mRNA metabolic process	SaOS-2	No	1.15×10^{-20}
mRNA metabolic process	SaOS-2	Yes	1.60×10^{-25}

Other recurring, and less expected top ten biological process GO terms included selenium-related GO terms (all groups except arsenite stressed U2OS) and 'SRP-dependent cotranslational targeting to membrane'.

'RNA binding' and 'poly(A) RNA binding' were the top two molecular function GO terms for all conditions. 'Adherens junction' was a top ten molecular function hit in stressed U2OS cells (adjusted p-value = 1.12×10^{-16}).

'Ribonucleoprotein complex' was the top cellular component GO term for U2OS (stressed and unstressed) and stressed HeLa. It was the second top hit for unstressed HeLa (after 'cytosolic small ribosomal subunit). It was the top hit among GO terms to which are annotated less than 1000 proteins in both SaOS-2 samples although it fell beneath other GO terms, such as 'extracellular exosome' (Bonferroni adjusted p-values = 1.56×10^{-20} [unstressed], 1.82×10^{-28} [stressed]).

From this chapter, it is claimed that HUWE1, SQSTM1, NARR and LDB1 are novel UNR-interacting proteins. It is further claimed that UNR mainly interacts proteins within ribonucleoprotein complexes, and that it may bind to proteins involved in selenium metabolism and at adherens junctions.

5 Identification of UNR-interacting transcripts

UNR is known to be an RNA binding protein, yet few UNR-interacting transcripts are currently recorded in the literature (Ray et al. 2015). Initial results in the lab had suggested that UNR may bind to a large number of transcripts in human cells (data not shown). Following on from those initial observations, this chapter set out to identify more UNR-interacting transcripts. The remaining 80% of the RIP samples generated for mass spectrometry (see sections 4.3, 4.5 and 4.6) were used to extract RNA that was then sequenced by RNA-Seq. The intention was to find RNAs that were present in statistically significantly higher quantities in the UNR immunoprecipitation samples than in the control IgG immunoprecipitation samples. Such RNAs would be expected to have be associated with UNR.

Dr Swagat Ray kindly carried out RNA extraction from the HeLa samples (section 2.2.13, steps 1-19) and partly carried out the RNA extraction from the U2OS samples (section 2.2.13, steps 1-12). Dr Emma Anderson kindly carried out the final resuspension and storage steps for the U2OS and SaOS-2 samples (section 2.2.13, steps 13-19). Lesley Ward of the University of Warwick Genomics Facility further processed the samples (see sections 2.2.15 and 5.1.1) before sending them to the Wellcome Trust for analysis (see section 5.1.1). Dr Nigel Dyer kindly pre-processed the raw RNA-Seq data that was received back from the Wellcome Trust (see section 5.1.1). He also allowed me to use and customise pieces of generic code that he had written to run PCA analysis in Matlab and DESeq2 analysis in R (see sections 2.2.18, 5.2 and 5.3).

5.1 RIP-RNA-Seq

5.1.1 Introduction to RNA-Seq

As stated previously, 80% of the RIP samples were set aside for the extraction and sequencing of UNR-associated RNA species (Section 4.1.1).

Briefly, a phenol/chloroform-based method was used to extract RNA from the immunoprecipitates directly from the beads. The RNA was then passed on to the University of Warwick genomics department who further processed the RNA using the low sample protocol

from the Illumina TruSeq RNA sample preparation v2 guide (see methods section), with the following modifications:

1) The oligo-dt purification step was omitted
2) Where possible, 100 ng of RNA was used per sample
3) Whereas the suggested bioanalyzer chip was used to verify RNA purity, the libraries were quantified using Qubit Fluorometry, which yields more precise results (Davies et al. 2016)

The libraries were normalised to 10 nM, made into two 10 nM pools and sent offsite to be sequenced by the Wellcome Trust facility in Oxford, using an Illumina HiSeq4000 on a 75 base pair paired end run. Illumina claim that their TruSeq RNA sample prep v2 fragmentation protocol generates fragments in the range of 120-200 base pairs with a median of 150 base pairs.

Whereas conventional DNA sequencing allows for direct mapping of reads to a reference genome, after accounting for intraspecific genetic variation, RNA sequencing has the additional problem of sequencing across splice junctions. This arises due to the process of splicing (Green 1986). Pre-messenger RNA (pre-mRNA) is transcribed from the antisense strand of DNA, resulting in a copy of the sense strand with thymine replaced with uracil. Pre-mRNA can contain introns that are spliced out as part of the process of generating mature mRNA. As part of the processing carried out by genomics, the first strand of complementary DNA (cDNA) was generated by reverse transcription of fragmented RNA, followed by the removal of the RNA template and the generation of a complementary strand of DNA. In the case of fragments of mRNAs that cross splice junctions, this can result in the generation of cDNA sequences that are not present in the genomic DNA from which they were ultimately generated (Figure 5.1).

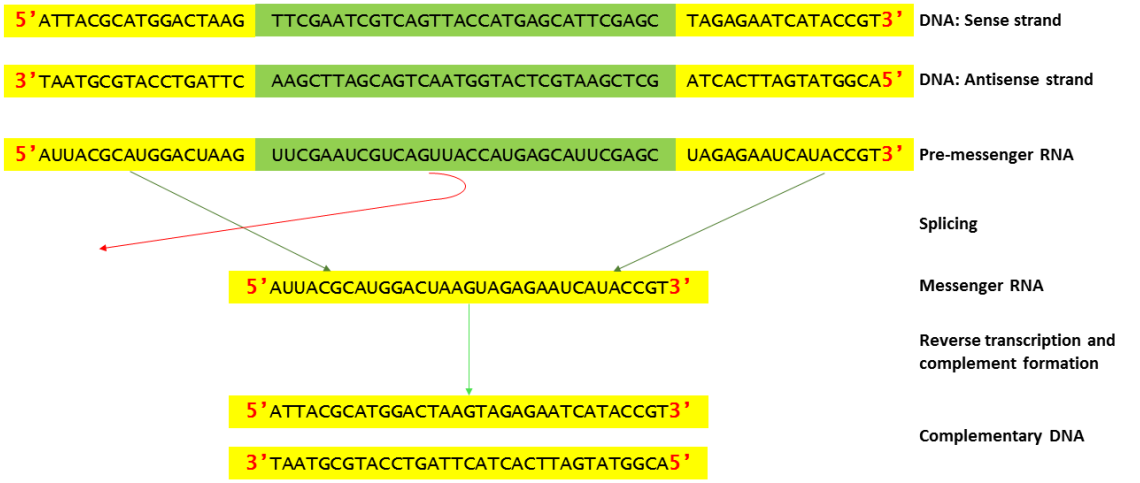


Figure 5.1: How cDNA differs from genomic DNA. Schematic diagram showing part of an mRNA fragment and how this is related to the DNA from which it is transcribed and to cDNA that can be generated from it.

Dr Nigel Dyer of the Ott lab kindly processed the raw data using Tophat2 and the HTSeq tool, htseq-count, before providing the author with raw count data within a Microsoft Excel file. Tophat2 is a program that uses Bowtie2 to map reads directly to a reference genome and this can deal with the problem of introns by locating regions of the genome that would align with a given sequence assuming that a single sequence was inserted into it (Kim et al. 2013; Langmead & Salzberg 2012). Htseq-count tallies unambiguous reads that overlap exons of specific genes (Anders et al. 2015). Raw count data at the gene level, as opposed to the transcript level, was thereby produced that was then exported to Microsoft Excel and passed on to the author for further analysis.

5.2 Principal component analysis on RNA-Seq data

Prior to proceeding with analysis of the RNA-Seq data, principal component analysis was first undertaken. Principal component analysis can be used to locate outliers and groupings within a group of samples (Wold et al. 1987). As it shows how the samples cluster by principal component, it is possible to confirm that known biological variability (e.g. cell type, drug treatment or whether the RIP was carried out with a test or control antibody, etc.) is responsible for the observed variation in the data.

Principal component analysis was carried out using Matlab, running a standard script used in the Ott lab that was kindly provided by Dr Nigel Dyer.

5.2.1 Initial round of PCA identifies 4 outliers in PC2 and/or PC1

The samples were identified by cell type, repeat number and drug treatment (i.e. arsenite stressed or unstressed), as with the mass spectrometry data. Due to potential confusion should the SaOS-2 cell type be followed by numbers, 'SaOS' was used as the cell name. How the numbers equate to condition is summarized below:

1, 2 and 3 = IgG pulldown repeats (1-3), unstressed,

4, 5 and 6 = UNR pulldown repeats (1-3), unstressed,

7, 8 and 9 = IgG pulldown repeats (1-3), arsenite stressed,

10, 11 and 12 = UNR pulldown repeats (1-3), arsenite stressed.

An initial round of PCA was carried out looking for outliers in the first two principal components, as these provide the greatest individual contributions to the overall variance in the data. Considering an outlier to be any sample that lies far away from the main cluster/clusters of samples, the first search showed that there were four clear outliers (Figure 5.2). As they lay far from the main cluster of samples, they contained a large source of variation of unknown origin that would confound subsequent analyses that included them. Those samples were therefore removed from subsequent rounds of PCA.

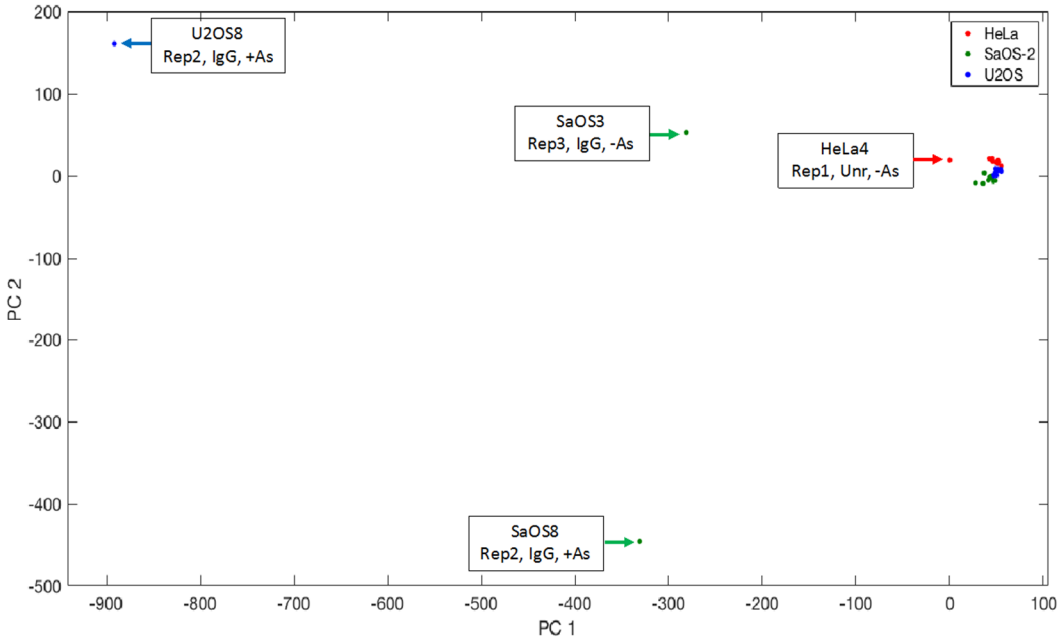


Figure 5.2: First round of PCA on RNA-Seq data. Red dots denote HeLa samples, green dots represent SaOS-2 samples and blue dots signify U2OS samples. The highlighted samples lay outside the general cluster and were rejected as outliers. They were removed from subsequent rounds of PCA.

5.2.2 Second round of PCA locates a further outlier in PC1 and PC2

Having removed the four outliers from the data, a second round of PCA was carried out on the remaining data, again looking for outliers in the first two principal components. Following the same reasoning as in section 5.2.1, this round of PCA highlighted one further outlier that was removed as with the previous four outliers (Figure 5.3).

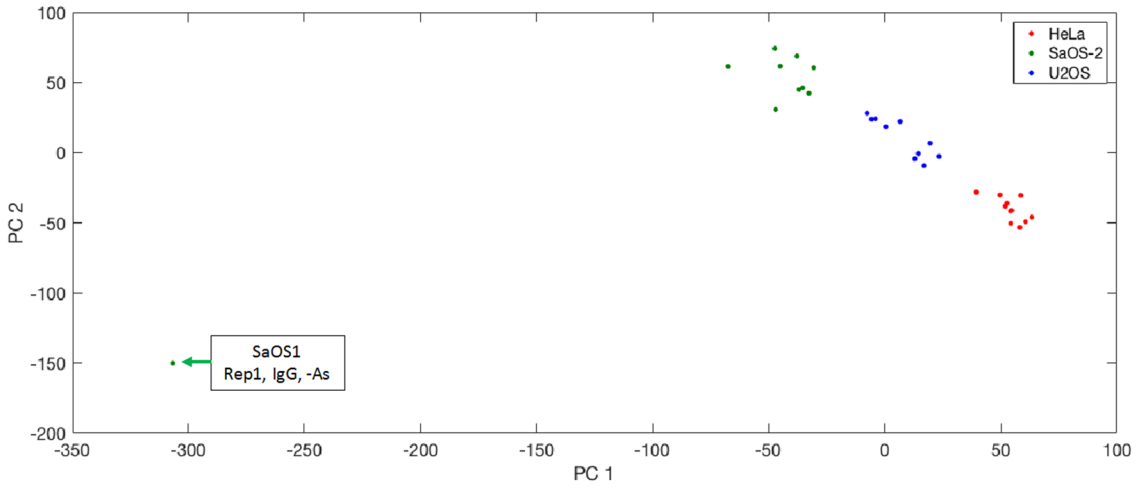
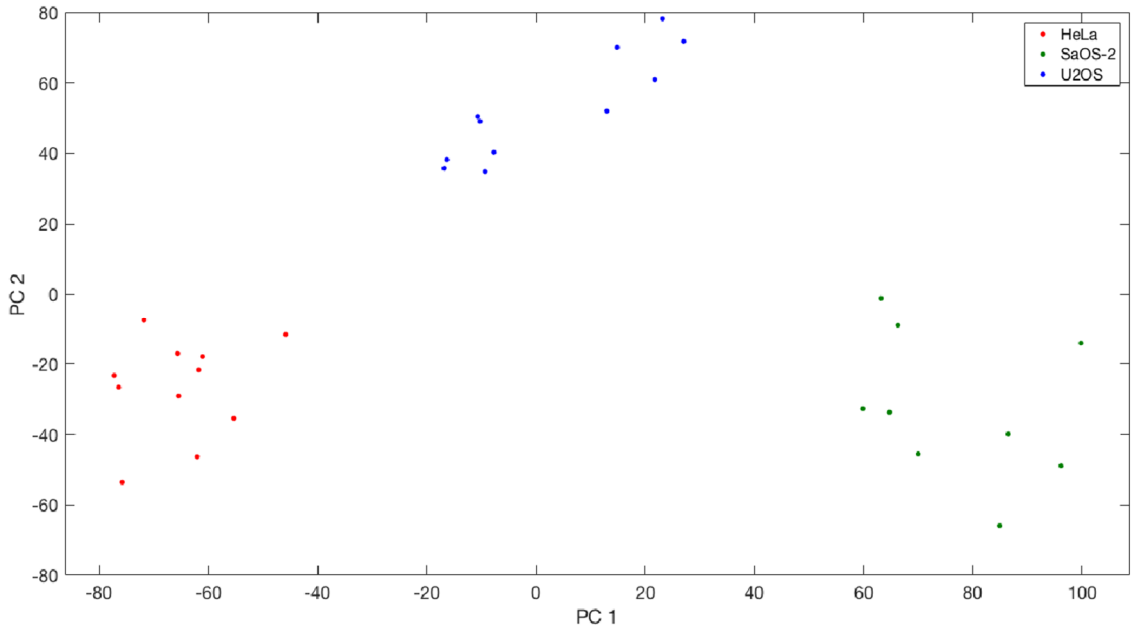


Figure 5.3: Second round of PCA on RNA-Seq data, having already removed four outliers. Red dots denote HeLa samples, green dots represent SaOS-2 samples and blue dots signify U2OS samples. The highlighted sample (SaOS-2 minus arsenite, IgG repeat 1) lies outside the general cluster and was rejected as an outlier from the subsequent round of PCA.

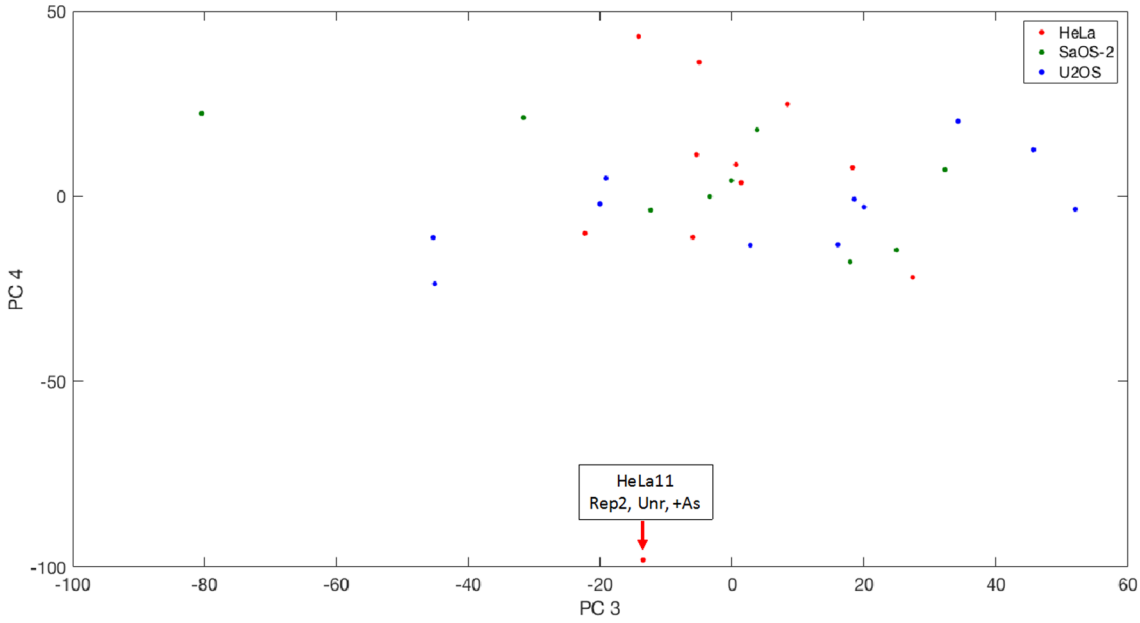
5.2.3 Third round of PCA locates no further outliers in the top two principal components

Having removed the fifth outlier, it was decided to look at whether or not the remaining samples clustered by cell type over the top six PCs (Figure 5.4).

A


HeLa
SaOS-2
U2OS
PC 2
PC 1
80
60
40
20
0
-20
-40
-60
-80
-80
-60
-40
-20
0
20
40
60
80
100

B


HeLa
SaOS-2
U2OS
HeLa11
Rep2, Unr, +As
PC 4
PC 3
50
0
-50
-100
-100
-80
-60
-40
-20
0
20
40
60

C

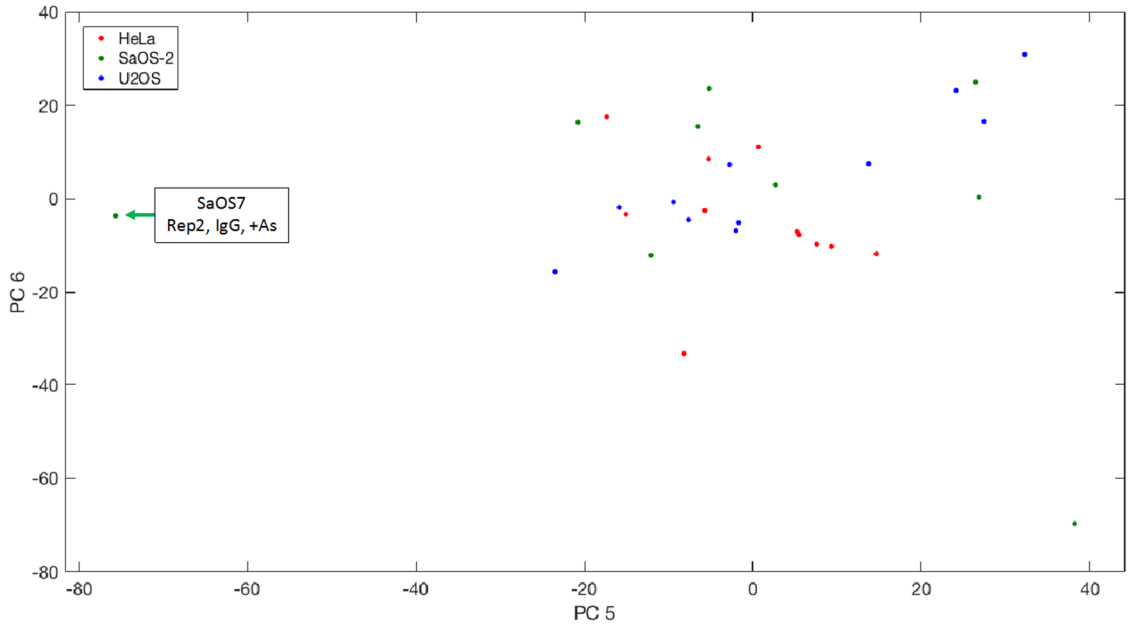


Figure 5.4: Third round of PCA on RNA-Seq data, having removed five outliers. Samples are coloured by cell type (HeLa = red, SaOS-2 = green and U2OS = blue) over PC1 and PC2 (Figure 5.4A), PC3 and PC4 (Figure 5.4B) and PC5 and PC6 (Figure 5.4C). The highlighted samples in Figures 5.4B and 5.4C lay outside the general cluster but were not rejected as outliers.

It was shown that the cell types were separated by PC1 (Figure 5.4A), implying that cell type accounts for the largest source of variation in the data. The *TP53*-wild type U2OS cell type was separated from the *TP53*-null SaOS-2 cell type and the *TP53*-disregulated HeLa cell type by PC2 (Figure 5.4A), although it is not fair to assume that *TP53* status was in any way responsible for this observation. None of PC3, PC4, PC5 nor PC6 appeared to account for variance due to differences in cell type as there was no clear clustering by cell type over these components (Figures 5.4B and 5.4C).

5.2.4 It was decided not to remove any further outliers from the third round of PCA

Two further potential outliers were flagged up in the third round of PCA; one in PC4 (Figure 5.4B) and one in PC5 (Figure 5.4C). It was decided not to remove these samples from the data as the variance accounted for by PC4 and PC5 is relatively small (Table 5.1). As a result of that, the

removal of more samples could not be justified. That is because, *ceteris paribus*, the removal of replicates reduces the ability of statistical tests to detect true differences and to reject false positives (Cohen 1992).

Table 5.1: The variance and cumulative variance accounted for by the first six principal components following the removal of 5 outliers.

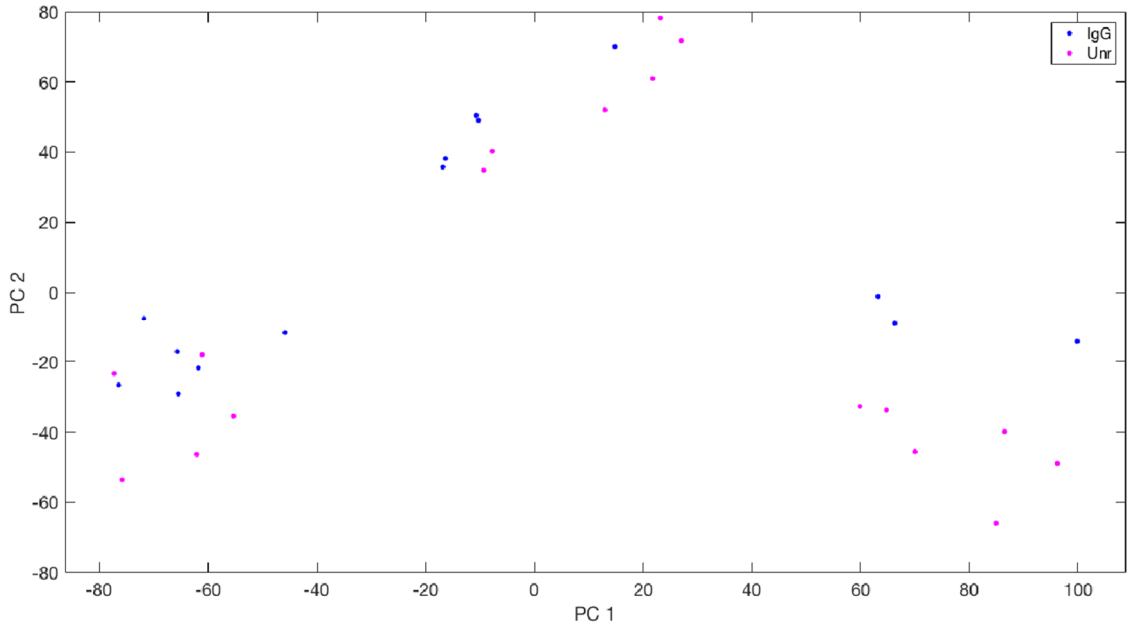
Principal component	Associated variance	Cumulative variance
PC1	34.0	34.0
PC2	17.9	51.9
PC3	7.6	59.5
PC4	5.7	65.3
PC5	4.5	69.7
PC6	3.5	73.3

It would take the sum of all principal components up to PC7 to attain a cumulative variance over 75% (76.1%), up to PC15 to reach 90% (90.5%) and up to PC20 to get to 95% (95.4%) (data not shown). Whilst the PCs lower than PC6 account for 26% of the variance in total, each individual component contributes relatively little to that total. Outliers confined to the lower principal components were therefore retained and principal components beneath PC6 were not examined graphically.

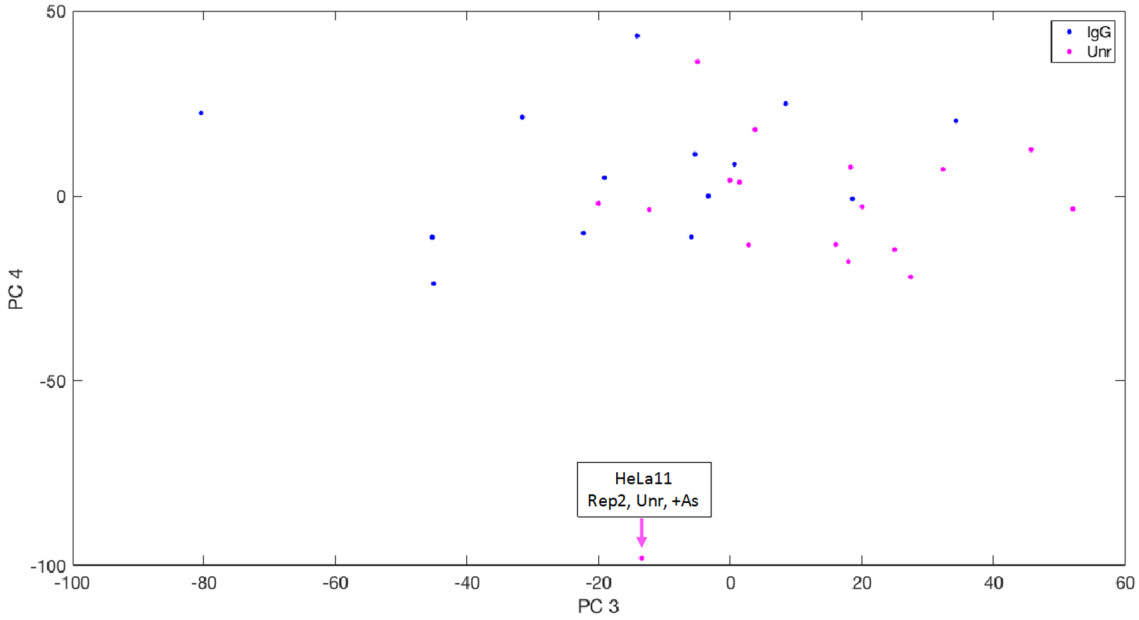
5.2.5 Third round of PCA shows effect of the choice of immunoprecipitating antibody is partly accounted for over the top three principal components

Having explored the way in which differences due to cell type are distributed over the top six principal components, it was decided to investigate whether or not some of the differences due to the choice of immunoprecipitating antibody (control IgG or anti-UNR) were also shared between the top six principal components (Figure 5.5).

A


IgG
Unr
PC 2
PC 1
80
60
40
20
0
-20
-40
-60
-80
-80
-60
-40
-20
0
20
40
60
80
100

B


IgG
Unr
HeLa11
Rep2, Unr, +As
PC 4
PC 3
50
0
-50
-100
-100
-80
-60
-40
-20
0
20
40
60

C

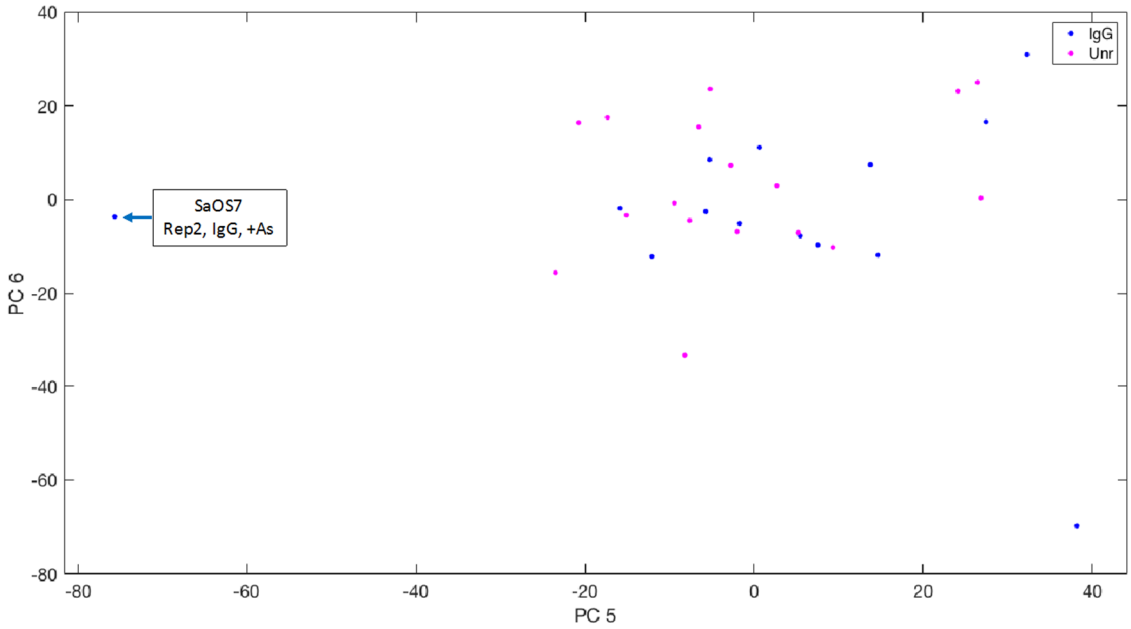


Figure 5.5: Third round of PCA on RNA-Seq data, having removed five outliers. Samples are coloured by immunoprecipitation antibody (control IgG = blue and anti-UNR = pink [termed 'Unr' in key]) over PC1 and 2 (Figure 5.5A), PC3 and 4 (Figure 5.5B) and PC5 and 6 (Figure 5.5C). The highlighted samples in Figures 5.5B and 5.5C lay outside the general cluster but were not rejected as outliers.

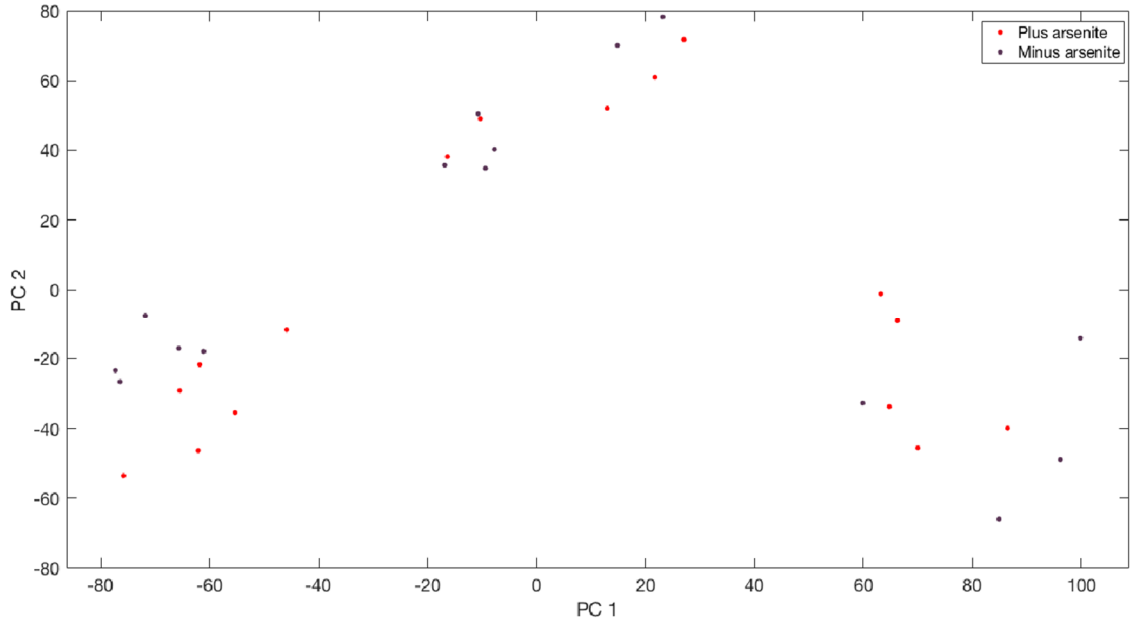
This suggested an interesting cell type dependent relationship over the first two principal components. In SaOS-2 (Figure 5.5A, bottom right cluster - see Figure 5.4A), the variance due to antibody choice appears to be in part explained by PC2. In U2OS (Figure 5.5A, top cluster - see Figure 5.4A), that variance seems to be shared between PC1 and PC2. In the HeLa samples (Figure 5.5A, bottom left cluster - see Figure 5.4A), the variance due to antibody choice does not seem to be so clearly explained by PC1 and/or PC2 as in the other cell types.

Principal component 3 appears to explain some of the variance due to the difference in antibody used in that most IgG samples tend to have values less than zero, whereas most UNR samples tend to have values greater than zero (Figure 5.5B). PC4, PC5 and PC6 do not appear to explain antibody-related variance (Figures 5.5B, 5.5C).

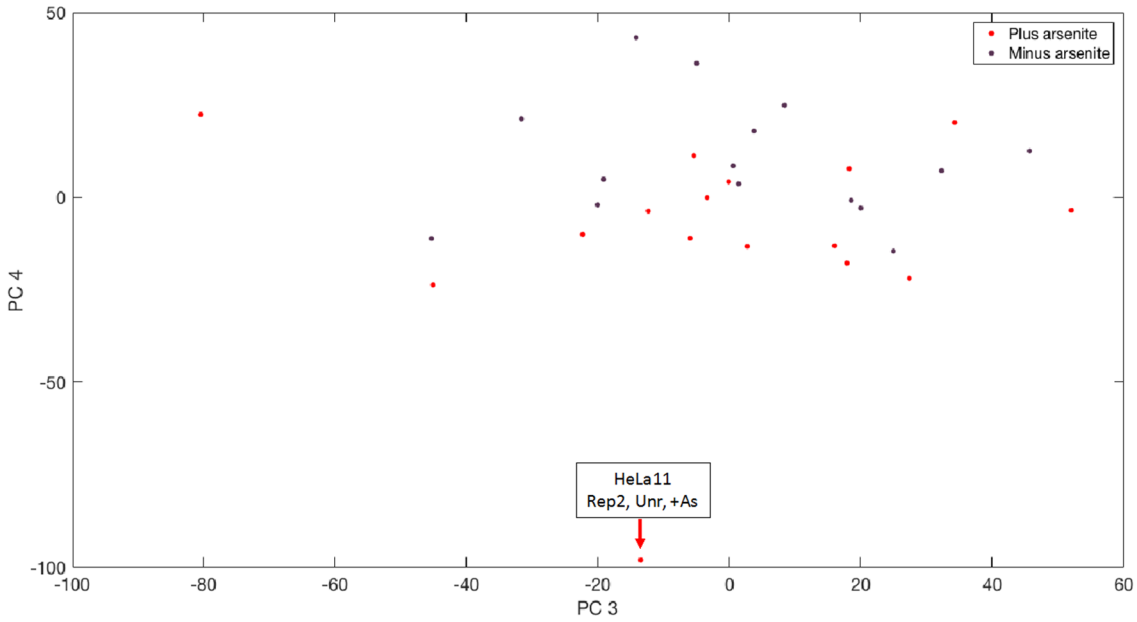
5.2.6 Third round of PCA shows effect of arsenite treatment is partly accounted for by PC1 and PC2 in HeLa samples and PC4 and PC5 more generally

Having looked at how the top six principal components can account for the variance due to cell type and immunoprecipitating antibody, it was considered prudent to look into how they could account for the other controlled variable in the experiment – arsenite treatment. This analysis is graphically displayed below (Figure 5.6).

A



B



C

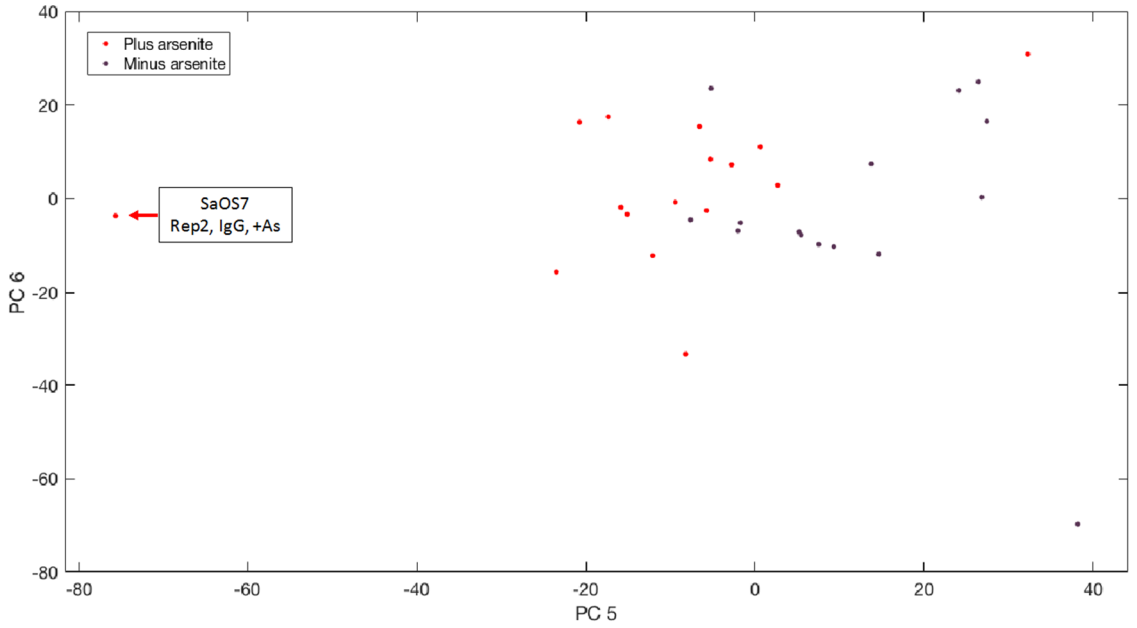


Figure 5.6: Third round of PCA on RNA-Seq data, having removed five outliers. Samples are coloured by whether or not the samples had been subjected to arsenite treatment (no arsenite treatment = purple and 1 mM arsenite for one hour immediately prior to harvesting = red) over PC1 and 2 (Figure 5.6A), PC3 and 4 (Figure 5.6B) and PC5 and 6 (Figure 5.6C). The highlighted samples in Figures 5.6B and 5.6C lay outside the general cluster but were not rejected as outliers.

There is some evidence for the variance due to arsenite treatment in HeLa being shared between PC1 and PC2, although this is less apparent for the other cell types (Figure 5.6A, compare Figure 5.4A for cell types). There is some evidence for PC4 (untreated samples generally more positive) and PC5 (untreated samples generally more positive) explaining some of the variability due to arsenite treatment, but this is not true for PC3 and PC6 (Figures 5.6B and 5.6C).

5.2.7 Summary of PCA work

Two rounds of PCA were carried out, during which five outliers were detected in the first two principal components. These outliers were the samples with the lowest total number of reads aligned to the human genome (Table 5.2).

Table 5.2: Total number of reads aligning to the human genome, by sample

Sample	Genome reads	%age of max		Sample	Genome reads	%age of max
HeLa1	15526162	54.0		U2OS1	16183300	56.3
HeLa2	15495709	53.9		U2OS2	19966464	69.5
HeLa3	12385277	43.1		U2OS3	20162372	70.1
HeLa4	720695	2.5		U2OS4	12252870	42.6
HeLa5	20301255	70.6		U2OS5	17327799	60.3
HeLa6	13866858	48.2		U2OS6	14690132	51.1
HeLa7	11878865	41.3		U2OS7	6517352	22.7
HeLa8	12375112	43.1		U2OS8	13011	0.0 (sic)
HeLa9	21143086	73.6		U2OS9	22790129	79.3
HeLa10	7051332	24.5		U2OS10	16715423	58.2
HeLa11	2409881	8.4		U2OS11	25932440	90.2
HeLa12	12996747	45.2		U2OS12	28743669	100
SaOS1	2235222	7.8		SaOS7	5379622	18.7
SaOS2	5605892	19.5		SaOS8	161914	0.6
SaOS3	110964	0.4		SaOS9	10224470	35.6
SaOS4	5203729	18.1		SaOS10	4509713	15.7
SaOS5	4627719	16.1		SaOS11	6952330	24.2
SaOS6	3486047	12.1		SaOS12	3690268	12.8

Key: Yellow = removed after first round of PCA (Figure 5.2), green = removed after second round of PCA (Figure 5.3), purple = the only other value less than 10% of maximum number of reads. This sample was one of the two samples suggested for removal after the third round of PCA (Figure 5.4B).

It has been shown that much of the expected variability due to the controlled experimental variables can be accounted for, at least in part, by the top 5 PCs. These components were furthermore calculated to account for around 70% of the total variation, implying that any expected variability accounted for by lower principal components would probably be small in

proportion to the total as each lower component accounts for an ever decreasing proportion of the overall variation.

As all of the controlled variables (cell type, arsenite treatment and immunoprecipitating antibody) had an effect on the samples, as visualized by PCA, the experimental approach was valid and there is likely to be useful information in the data.

Having decided upon which of the samples to remove as outliers, it was decided to proceed and to analyse the data further using DESeq2.

5.3 DESeq2

DESeq2 is an R-based bioinformatics tool that can analyse count data, such as RNA-Seq data, and provide information about differential expression (Love et al. 2014). It is worth pointing out that the interesting comparisons between these data are not strictly questions of differential expression, but of differential immunoprecipitation.

DESeq2 was compared against a variety of similar programs in the Ott lab by Dr Nigel Dyer, who concluded that DESeq2 was the most consistent and reliable program (personal correspondence). Other work in the literature has also suggested that DESeq2 is a good choice of program for differential expression (Ching et al. 2014).

5.3.1 Summary of samples remaining following rejection of five samples following PCA

The RIP-Seq experiment had 6 conditions – plus and minus arsenite for each of the three cell types. There were three UNR pulldown samples for each condition, each of which had a paired IgG sample. There were three repeats, thereby making 36 samples in total. It had been decided following principal component analysis that five samples were to be treated as outliers and

removed (see section 5.2.6). The total number of samples remaining following the rejection of suggested outliers was, therefore, 31.

Each cell type/arsenite treatment condition started with three pairs (i.e. the three repeats). Following outlier removal, only HeLa plus arsenite and U2OS minus arsenite retained all three UNR/IgG pairs (Tables 5.2A and 5.2B). SaOS-2 minus arsenite, on the other hand, retained only one IgG sample but all three UNR samples (Table 5.3).

As no UNR/IgG pairs were both designated as outliers, five of the eighteen possible pairs were lost. HeLa plus arsenite and U2OS minus arsenite retained 3 pairs, SaOS-2 minus arsenite retained one pair and the other three conditions had retained two pairs each (Table 5.4).

Table 5.3: Remaining samples following the removal of outliers

Cell type	- arsenite IgG	- arsenite UNR	+ arsenite IgG	+ arsenite UNR
HeLa	H1, H2, H3	H5, H6	H7, H8, H9	H10,H11, H12
U2OS	U1, U2, U3	U4, U5, U6	U7, U9	U10, U11, U12
SaOS-2	S2	S4, S5, S6	S7,S9	S10, S11, S12

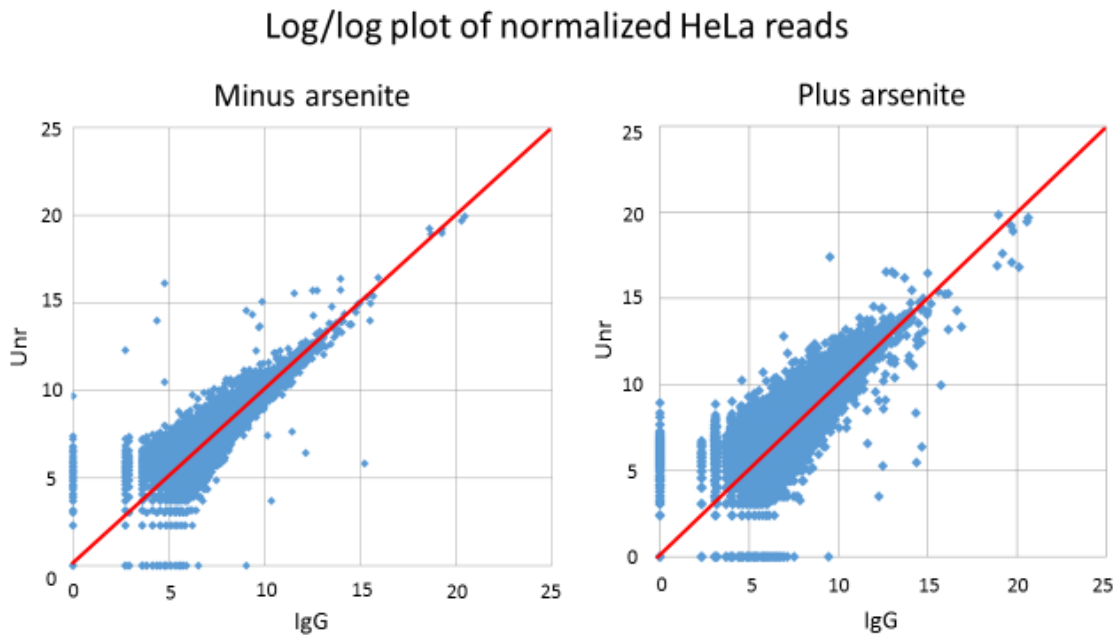
Table 5.4: Remaining pairs following the removal of outliers and their paired samples

	Remaining pairs	
Cell type	- arsenite	+ arsenite
HeLa	H2/5, H3/6	H7/10, H8/11, H9/12
U2OS	U1/4, U2/5, U3/6	U7/10, U9/12
SaOS-2	S2/5	S7/10, S9/12

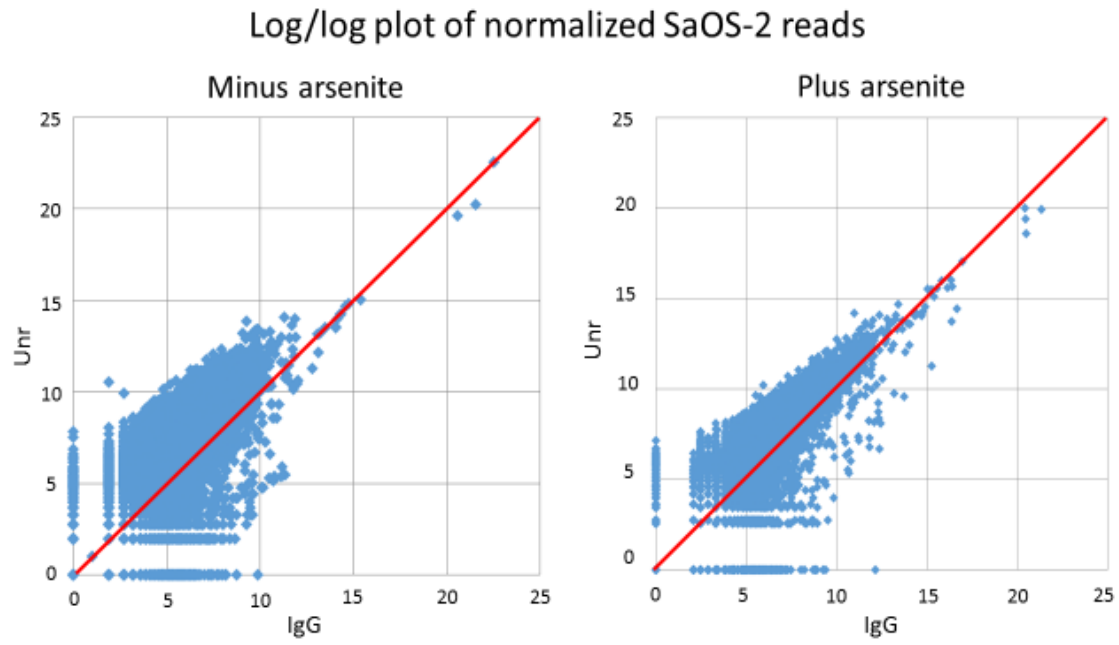
In order to look at the retained data graphically, all the samples in Table 5.4 were normalized by their total read counts and arranged in UNR/IgG pairs by gene. These data then had 1 added to every count value prior to having their log2 value calculated. The reason for adding 1 was to avoid the risk of having to take logs of values less than one and potentially zero, which would make the point unplottable. By adding 1, every log value had to be greater than or equal to 0.

Every retained pair of UNR and IgG samples for each condition then had their log2 values plotted for every gene, thereby generating six plots (Figure 5.7). As the data had been normalized, the total number of adjusted reads for each sample were the same. Nevertheless, there appeared to be a general trend for points (i.e. gene observations) to lie further above the line y=x than to lie under it. Due to normalization, it is expected that this is explained by there being more individual points under the line than over it but the high concentration of points mean that this is not visible. Such a trend does imply that UNR is binding to some RNAs specifically. Individual points that lay far from the cluster around y=x were expected to be more significant than those that lay closer to the line y=x. Such graphs are plotted below by cell type and arsenite treatment (Figure 5.7).

A


Log/log plot of normalized HeLa reads
Minus arsenite
Plus arsenite
Unr
IgG
0
5
10
15
20
25

B


Log/log plot of normalized SaOS-2 reads
Minus arsenite
Plus arsenite
Unr
IgG
0
5
10
15
20
25

C

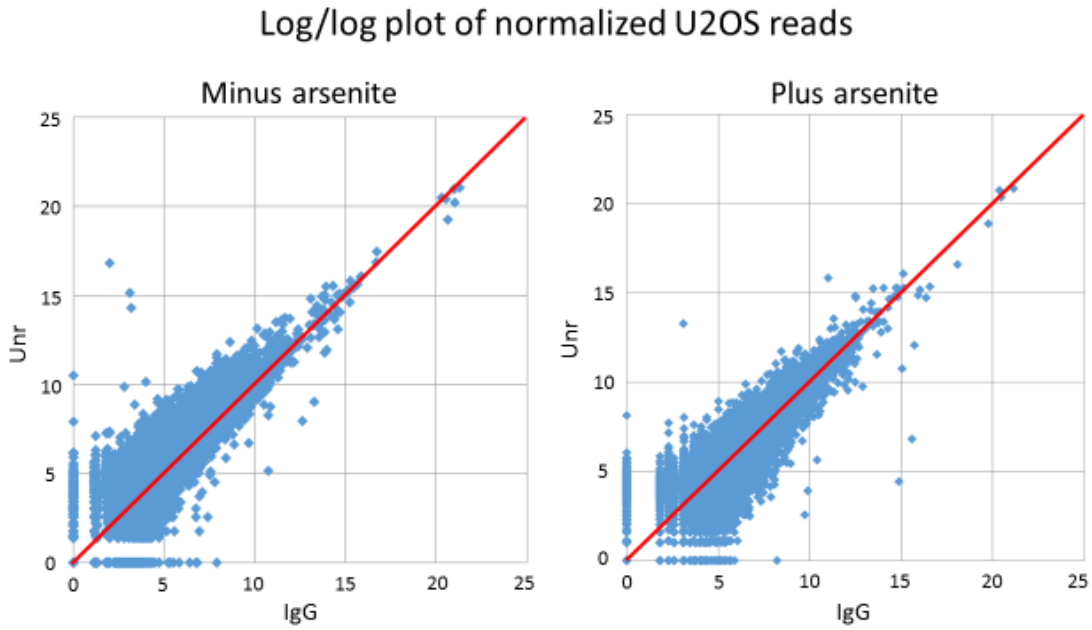


Figure 5.7: Log/log plots of normalized UNR (termed 'Unr') and IgG read count pairs for each gene, by cell type (A = HeLa, B= SaOS-2, C = U2OS) and arsenite treatment (as stated). Points refer to genes from the respective UNR/IgG paired samples in Table 5.4 (therefore, other than for SaOS-2 minus arsenite, every gene is present more than once). Briefly, 1 was added to each read count and the log2 of these values was taken and used to plot the graphs. Points lying on the red y=x line refer to RNAs that bound equally to UNR and IgG in a given repeat.

In general, the log transformation plots showed that there was a slight general skew towards more binding in the UNR samples. It is difficult to infer much about significance from these plots but it does look as though HeLa minus arsenite, for example, may have some genes that could be overrepresented in UNR at moderate read counts (log values 13-16) and a smaller number overrepresented in IgG (log value 12-15) (Figure 5.7A). Only following analysis using DESeq2 will it be apparent if there are any significant hits.

5.3.2 Strategy for using DESeq2 analysis to look for UNR-associated RNAs

It had been decided to use DESeq2 to look for differences between the RNAs pulled down by UNR and IgG for different groups of samples. The data were fed into DESeq2 for one factor

analysis (i.e. UNR for one condition being compared against IgG for the same condition) or, when considered appropriate, two factor analysis. In this case, the two factors would be:

1) Differences due to immunoprecipitating antibody,

2) Differences due to date of sample preparation.

Every cell type was processed individually and, therefore, the maximum number of members of each of the second factor groups was four – two pairs (UNR or IgG pulldown using lysates formed from stressed or unstressed cell harvested on a given day).

The choice of using one or two factor analysis was decided based upon the question that each method addresses. One factor analysis simply compares overall mean values for UNR and IgG. It therefore addresses the question of whether or not a given RNA bound preferentially and consistently in UNR pulldowns over IgG pulldowns.

As cell type accounted for the largest portion of variation in the data, comparing sample means across cell types could miss trends in preferential binding that could be skewed by differential expression levels between cell types. Where only one cell type was considered, one factor analysis was considered preferable as demonstrating a clear difference between means is intuitively pleasing.

The two factor analysis addresses a different question – whether there is a consistent difference between UNR and IgG pulldowns using lysates made on the same day. That consideration made it more appropriate for investigating UNR-interacting RNAs across cell types. Under such circumstances, it may find more UNR-interacting RNAs but the individual hits would be confounded with variation in expression levels. It could also generate results that are confounded with arsenite treatment as some cell type repeats have four samples (UNR and IgG, plus and minus arsenite) and some only have two (UNR and IgG, plus or minus arsenite).

To summarise, one factor analysis would be able to locate a subset of true hits that did not bind to IgG at any appreciable level (Figure 5.8A) or that were expressed at similar levels in all cell

types (Figure 5.8B) but two factor analysis would be required to locate true hits where expression varied greatly between, or within, cell types (Figure 5.8C).

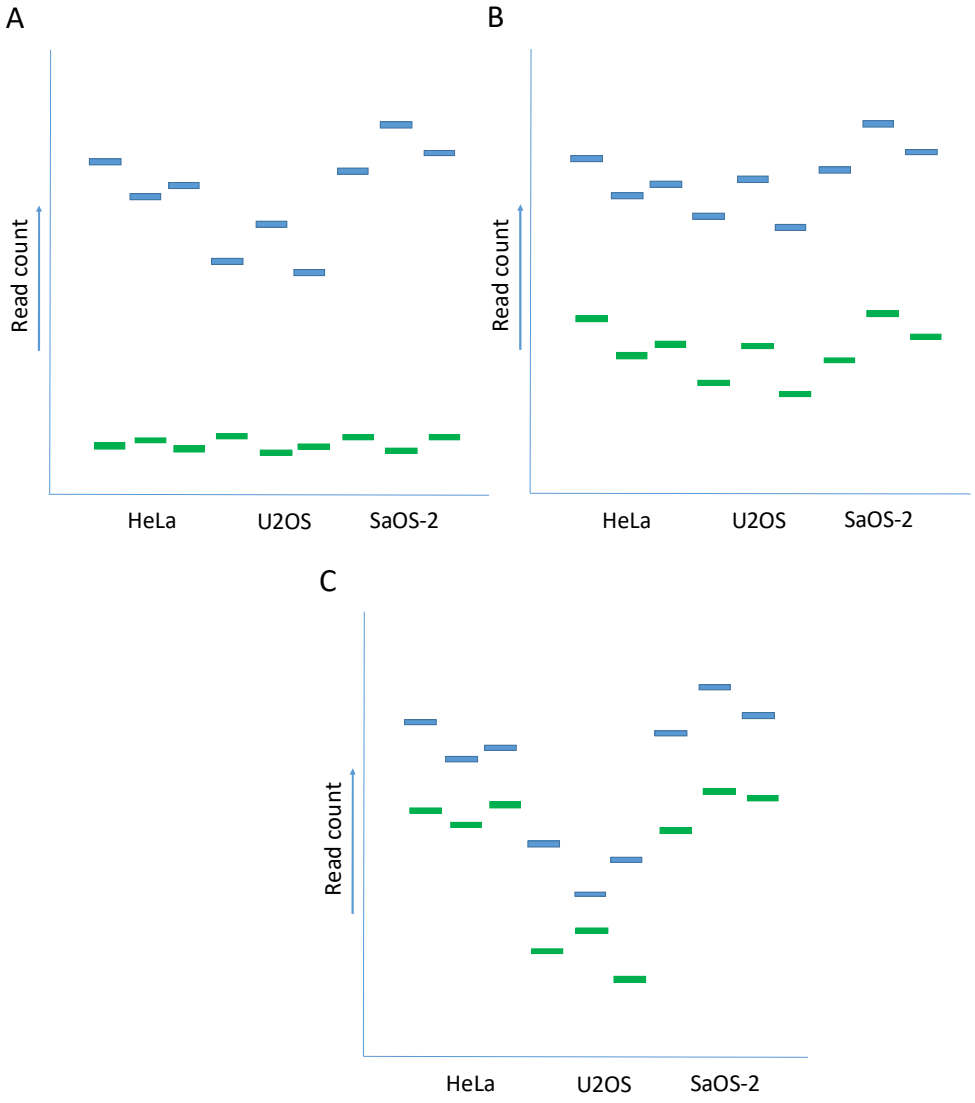


Figure 5.8: Schematic representation of three scenarios in which a hypothetical RNA binds preferentially to UNR (blue) over IgG (green) in three repeats of three cell types (HeLa, U2OS and SaOS-2). The scenarios show (A) a situation in which there is no appreciable binding of the RNA to IgG, (B) a situation in which there are similar levels of expression in all three cell types, (C) a situation in which the expression level varies between the cell types.

As two factor analysis compared results from UNR and IgG samples prepared on the same day, it was decided to reject all unpaired samples and only the pairs documented in Table 5.4 were

retained for further consideration by two factor analysis. As results obtained by two factor analysis and one factor analysis were to be compared and contrasted, it was considered helpful to also exclude unpaired samples from consideration by one factor analysis.

It was decided to compare samples on four levels:

1) By cell type and arsenite treatment
2) By cell type only (i.e. arsenite treatment independent)
3) By arsenite treatment only (i.e. cell type independent)
4) All samples combined (i.e. arsenite treatment and cell type independent)

By only providing DESeq2 with a subset of the samples, variation present in other groups of samples could not affect the ability of the program to detect differences that were specific to the subset (e.g. the HeLa plus arsenite samples would not be merged with HeLa minus arsenite nor samples from other cell types that could have a different distribution of read counts).

A problem with working with small groups of samples is that they are more affected by outliers. In this work, PCA analysis was used to highlight and remove outliers but the final round of analysis suggested further outliers that were retained for the reasons explained previously. On top of this, small groups of samples could be unrepresentative of a parent population containing all possible samples. It has been suggested that the majority of all biomedical research findings are wrong because of reasons such as small sample sizes (Ioannidis 2005). In order to increase the number of repeats for statistical testing using DESeq2, it was decided to merge similar groups. Firstly, as the PCA analysis suggested that the majority of the variability in the data was due to cell type differences rather than arsenite treatment, it was decided to merge the plus and minus arsenite samples for each cell type. This would have the positive effect of increasing the number of replicates for each condition, the condition always being the counts from UNR or IgG pull downs. On the other hand, the effect of arsenite would become a confounding variable. Where an RNA binds well to UNR under arsenite stress but poorly otherwise, merging the samples would reduce the significance of such an RNA. These mergers, therefore, look primarily for RNAs that bind preferentially to UNR irrespective of whether or not the cells from which they were obtained had been treated with arsenite or not.

A second merger went by arsenite treatment rather than cell type. This merger was looking for RNAs that bound to UNR across more than one cell type but that may have been affected by arsenite treatment. As discussed below, two factor DESeq2 analysis was considered more appropriate for analyzing these data.

Finally, all samples were merged together. This provided the greatest number of repeats but was confounded with both cell type differences and arsenite treatment. It was hoped that two factor DESeq2 analysis could find some UNR-associated RNAs that interact independently of cell type and arsenite treatment. Findings from this grouping and the previous merged groups were considered most likely to give information about any highly conserved roles for UNR in human cells.

5.3.3 A brief word on nomenclature

RNAs were pulled down in the RIPs and these were converted to cDNA for sequencing. Pre-processing of the sequencing data by Dr Nigel Dyer produced a list of non-transcript specific genes that were then used for further analysis (e.g. GO term analysis). In general, only protein coding genes were considered. Where protein coding genes were considered, this implied that their transcripts were pulled down in the RIP. Nevertheless, the data referred to genes and the term 'gene' will be used to describe the pulled down RNAs. Likewise, consideration will be made as to the function of proteins encoded by the significant genes. It should always be borne in mind that the original RIPs pulled down RNAs for sequencing and that consideration of proteins and the use of the word 'gene' in analytical discussion of the sequencing data should be understood in that context.

Where required, recommended protein names were used, as per www.uniprot.org, as well as the short version of gene names (associated gene names) for the GRCh38.p5 version of the human genome that was previously available for download via http://www.ensembl.org/biomart/martview/.

5.3.4 Technical details for gene ontology (GO) term overrepresentation analyses

Unless specified otherwise, the technical details of the GO term searches carried out for this chapter, as obtained from the GO tool, are presented in Table 5.5.

Table 5.5: Technical information associated with GO term searches

Analysis Type:	PANTHER Overrepresentation Test (release 20160715)
Annotation Version and Release Date:	GO Ontology database Released 2016-06-22
Analyzed List:	upload_1 (Homo sapiens)
Reference List:	Homo sapiens (all genes in database)
Bonferroni correction:	TRUE

5.4 DESeq2 results for HeLa samples

5.4.1 Results comparing HeLa minus arsenite samples by immunoprecipitating antibody

The two pairs of HeLa minus samples (Table 5.4) were compared by whether they were from UNR or IgG pulldowns. DESeq2 only highlighted one gene as having a Benjamini-Hochberg (BH) adjusted p-value less than 0.05 – *ZNF8* (adjusted p=0.0397). This gene was higher in the UNR samples and was protein coding.

Zinc finger proteins contain the zinc finger DNA binding motif (Berg 1990). Znf8 (Zinc Finger Protein 8) is a repressor of expression as a result of Tgf-β/Bmp signaling (Jiao et al. 2002). This is interesting as the UNR-interacting protein (UNRIP) is also involved in TGF-β signaling and this opens the possibility that UNR could influence signaling through protein-protein interactions as well as protein-transcript interactions.

Whilst it was decided to use one factor DESeq2 analysis to generate results that are easier to interpret, it was decided to carry out two factor analysis on the HeLa minus arsenite samples. This changed the comparison from one in which all UNR samples should be higher than all IgG samples to one in which all comparisons should bind more to UNR than to IgG. That can be considered a positive change in one respect as the idea of these analyses was to find RNA species that are pulled down preferentially in UNR pull downs over over IgG pull downs. Also, it is more difficult to interpret plots of the results of two factor analyses as mean read counts may be skewed up or down by the absolute read count whereas the comparison was carried out on the difference between pairs of read counts. Nevertheless, it was considered useful to carry out a two factor analysis for the HeLa minus arsenite samples to see if it generated more hits and, if so, if they could be used to generate significant GO terms.

The two factor analysis generated 42 genes with a BH-adjusted p-value less than 0.05, of which 41 were higher in the UNR samples, of which 37 were protein coding. *ZNF8* remained a hit (new adjusted p-value = 0.01722) but the new top hit was *EIF2AK2* (Interferon-induced, double-stranded RNA-activated protein kinase) which had an adjusted p-value of 1.39×10^{-11}. EIF2AK2 is a proapoptotic factor (Lee & Esteban 1994) as well as acting to shut off translation in response to infection with certain viruses (Garcia et al. 2006). More recently, it has also been implicated in Alzheimer's disease (Bullido et al. 2008). Interestingly, UNR has been associated with apoptosis, viral translation and Alzheimer's disease (all discussed in chapter 1).

Entering the 37 protein coding genes into the GO term enrichment tool at http://amigo.geneontology.org/rte generated no significant Bonferroni-adjusted GO terms (biological process, molecular function and cellular component).

In the absence of significant GO terms, the protein coding genes are tabulated below with their DESeq2-calculated UNR/IgG fold changes and BH-adjusted p-values (Table 5.6).

Table 5.6: Putative UNR-interacting transcripts from the HeLa minus arsenite samples with two factor DESeq2 analysis-calculated UNR:IgG ratios and BH-adjusted p-values

Gene	UNR/IgG ratio	BH-adjusted p-value	Gene	UNR/IgG ratio	BH-adjusted p-value
EIF2AK2	3.95	1.39E-11	*TIMM50*	2.27	0.014295
LYRM7	3.59	1.67E-05	*ZNF8*	2.55	0.017221
XIAP	3.14	5.68E-05	*NDUFV3*	2.38	0.017546
MRI1	2.81	0.000127	*AHR*	2.35	0.017546
VHL	3.02	0.000172	*ZNF714*	2.45	0.022859
ZNF587	3.14	0.000259	*GNB4*	2.07	0.031744
CYP20A1	3.03	0.000608	*MCM8*	2.39	0.031785
LIMD1	2.63	0.000828	*ATP5S*	2.43	0.032897
BRI3BP	2.45	0.000828	*BPNT1*	2.36	0.032897
MDM4	2.51	0.001199	*YIPF4*	2.27	0.032897
DFFA	2.37	0.002511	*SPPL2A*	2.17	0.032897
CCNYL1	2.61	0.002634	*EMP2*	1.95	0.036506
MMACHC	2.57	0.005900	*MIER1*	2.10	0.036727
METTL2B	2.54	0.005900	*DTWD2*	2.35	0.044065
PSMD12	2.33	0.005900	*ZNF556*	2.36	0.044687
APOOL	2.59	0.009157	*TMOD3*	1.98	0.044687
MYLK3	2.22	0.011044	*POLR2D*	2.10	0.047449
MPV17L	2.62	0.012438	*CRCP*	2.15	0.049507
FAM20B	2.11	0.013412			

5.4.2 Results comparing HeLa plus arsenite samples by immunoprecipitating antibody

The three pairs of HeLa plus samples (Table 5.4) were then compared by whether they were from UNR or IgG pulldowns. DESeq2 highlighted 65 genes as having a Benjamini-Hochberg (BH) adjusted p-value less than 0.05 of which 51 were higher in the UNR samples. Of the 51 genes

that were significantly higher in the UNR samples, 46 were protein coding. The significant protein coding genes are presented on a log/log graph below (Figure 5.9). Normalized read counts from two HeLa plus arsenite UNR pulldown samples are plotted to give an idea of the spread of count read due to biological variation between samples. The values plotted were log2 values of the read counts for each condition, as stated on the graph axes, plus one. The graph also shows the average values for the significant genes. These were calculated by taking a mean value for the UNR and IgG counts of each gene, adding one and taking the log2 value of the result. The resultant UNR and IgG values were plotted, as stated on the axis labels.

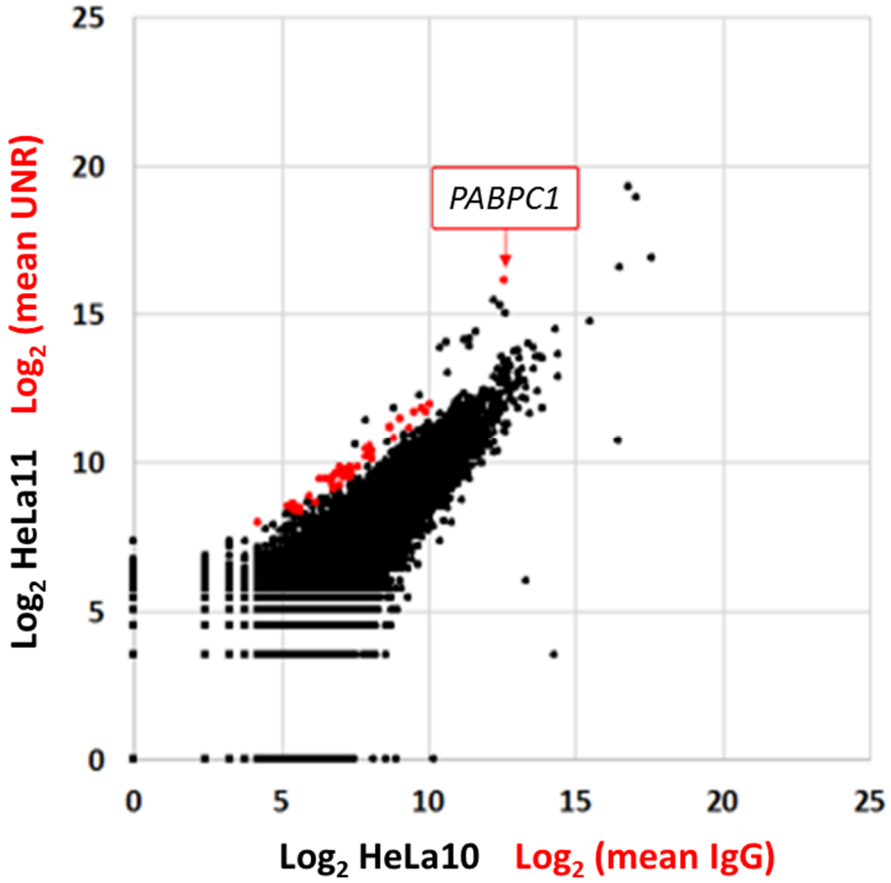


Figure 5.9: Log/log plot of one factor DESeq2 hits (red) and variability between two representative UNR pulldown samples (black) for HeLa plus arsenite. The name and location within the graph of the top hit is highlighted.

The top hit was *PABPC1* (BH adjusted p-value = 2.52×10^{-12}). It has already been well established that UNR interacts with both the mRNA and protein products of the *PABP* gene, both in the Anderson lab and in the literature (Patel et al. 2005; Chang et al. 2004).

With the exception of PABPC1, the significant hits had small UNR/IgG fold changes and were within the range of variation between HeLa10 and HeLa11 (Figure 5.9). This was in keeping with the PCA analysis that suggested that the immunoprecipitating antibody was a relatively minor contributor to the overall variation in the data.

The 46 protein coding genes were entered into the GO tool which returned no Bonferroni-multiple testing correction-adjusted molecular function nor cellular component GO terms. It gave 6 significant biological process GO terms (Table 5.7).

Table 5.7: Over-represented biological process GO terms generated using putative UNR-interacting protein coding genes designated as significant at a BH-adjusted p-value cut off of 0.05 using one factor DESeq2 analysis of arsenite stressed HeLa samples.

GO biological process complete	Total	Observed	Expected	Fold change	Adjusted p-value
heterocycle metabolic process (GO:0046483)	4620	25	10.13	2.47	1.46E-02
cellular aromatic compound metabolic process (GO:0006725)	4669	25	10.24	2.44	1.79E-02
aromatic compound biosynthetic process (GO:0019438)	3085	20	6.77	2.96	2.05E-02
nucleobase-containing compound metabolic process (GO:0006139)	4484	24	9.84	2.44	3.48E-02
organic cyclic compound biosynthetic process (GO:1901362)	3193	20	7	2.86	3.53E-02
organic cyclic compound metabolic process (GO:1901360)	4856	25	10.65	2.35	3.80E-02

In total, 16 gene were considered significant by two factor analysis of the HeLa minus arsenite data and by one factor analysis of the HeLa plus arsenite data (*AHR, ATP5S, BPNT1, CRCP, CYP20A1, EIF2AK2, LIMD1, LYRM7, MDM4, METTL2B, MMACHC, MPV17L, MRI1, TIMM50, ZNF556, ZNF587*).

5.4.3 Results comparing all HeLa samples by immunoprecipitating antibody

It was then decided to merge all the HeLa samples together for analysis. Of the five pairs of samples, 40% were non-arsenite treated and 60% were arsenite treated (Table 5.4).

One factor DESeq2 analysis of the merged data resulted in 132 BH-adjusted significant genes. Of these, 124 were higher in UNR, of which 111 were protein coding. *PABPC1* was the top hit from

the merged data, with an adjusted p-value of 2.36×10^{-14}. A log/log graph showing the significant protein coding genes is presented below (Figure 5.10).

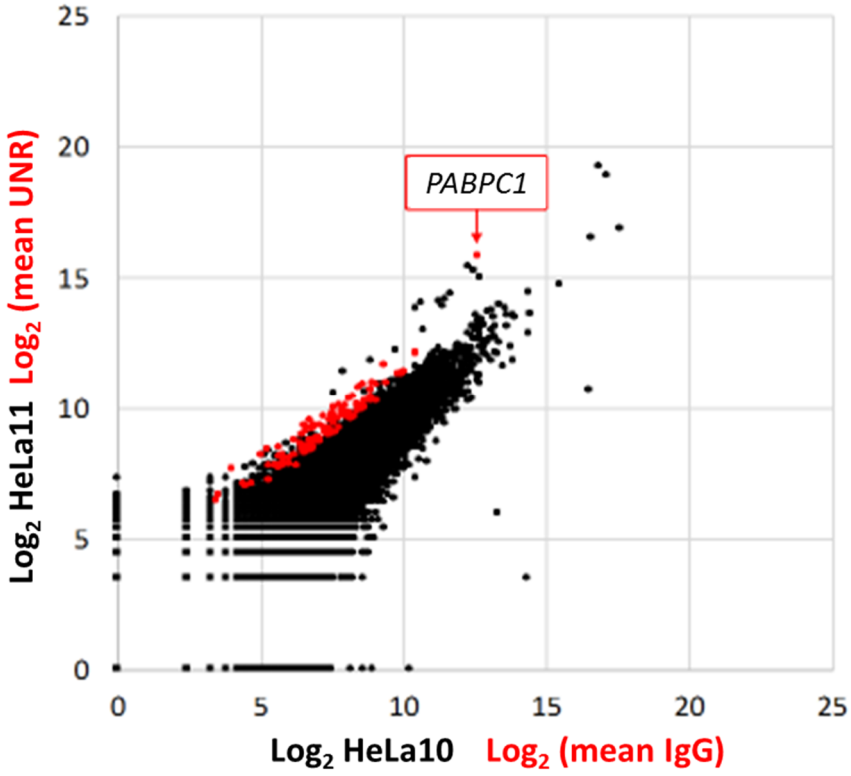


Figure 5.10: Log/log plot of one factor DESeq2 hits (red) and variability between two representative UNR pulldown samples (black) for all HeLa samples. The name and location within the graph of the top hit is highlighted.

In keeping with the PCA results and the observations made for the HeLa plus arsenite one factor analysis, the fold changes for the significant genes were small. A large number of the significant hits lay within the main body of the HeLa10/HeLa11 plot (Figure 5.10). The same samples were used as representative of all HeLa samples as the black dots would be exactly the same and the change in the location of the red dots could be viewed in relation to them. As could be seen, the significant hits using all HeLa samples increased in number relative to those for HeLa plus arsenite samples alone and expanded into the main body of the black dots.

The 111 significant protein coding genes that were higher in UNR were submitted to the GO tool which could not recognise two of them, resulting in 109 genes being considered in total. This found four related significant biological process GO terms but no significant molecular function nor cellular component GO terms (Table 5.8).

Table 5.8: Over-represented biological process GO terms generated using putative UNR-interacting protein coding genes designated as significant at a BH-adjusted p-value cut off of 0.05 using one factor DESeq2 analysis of all HeLa samples.

GO biological process complete	Total	Observed	Expected	Fold change	Adjusted p-value
cellular nitrogen compound metabolic process (GO:0034641)	5090	49	26.45	1.85	1.55E-02
heterocycle metabolic process (GO:0046483)	4620	46	24.01	1.92	1.58E-02
nucleobase-containing compound metabolic process (GO:0006139)	4484	45	23.31	1.93	1.78E-02
cellular aromatic compound metabolic process (GO:0006725)	4669	46	24.27	1.9	2.14E-02

5.4.4 Summary of the DESeq2 analyses using HeLa samples

It was noted that there were more putative UNR-interacting transcripts detected by DESeq2 in the HeLa plus arsenite samples than in the HeLa minus arsenite samples. Initial biological process GO term analysis suggested that UNR may interact with transcripts that encode proteins involved in such things as “nucleobase-containing compound metabolic processes”, at least in HeLa cells where at least 60% of samples had been stressed with arsenite.

5.5 DESeq2 results for U2OS samples

5.5.1 Results comparing U2OS minus arsenite samples by immunoprecipitating antibody

U2OS minus arsenite was one of the two conditions in which there were no outliers removed, the other being HeLa plus arsenite (Table 5.4). DESeq2 analysis of the U2OS minus arsenite samples gave 3 hits with a BH-adjusted p-value under 0.05, all of which were higher in the UNR samples. Only one of the three was protein coding, Interleukin-8 *CXCL8* (adjusted p-value = 0.0022).

CXCL8 is involved in various lung pathologies (Mukaida et al. 2003). One biological function of CXCL8 is to induce angiogenesis, which it carries out by interaction with CXCR2 (Heidemann et al. 2003). As well as a potentially pro-oncogenic role in stimulating angiogenesis, CXCL8 has also been shown to promote metastatic events in colorectal cancer (Xiao et al. 2015).

As there was only one gene found by the one factor analysis, it was decided to carry out two factor analysis on the U2OS minus arsenite samples. This resulted in the loss of the single protein coding hit (*CXCL8*). In summary, five genes were significant at a BH-adjusted p-value cut off of 0.05, of which 3 were higher in the UNR samples but none of these three were protein coding.

5.5.2 Results comparing U2OS plus arsenite samples by immunoprecipitating antibody

One factor DESeq2 failed to locate any significant hits from among the genes present in the two pairs of U2OS plus arsenite samples (Table 5.4). Two factor analysis located 75 hits, 7 of which were both protein coding and higher in the UNR samples. *PRR11* had the most significant BH-adjusted p-value (0.0011). The seven genes did not yield any significant GO terms. The seven genes are tabulated in Table 5.9. PRR11 has been associated with epithelial-mesenchymal transition (EMT) and, thereby, metastasis in breast cancer (Zhou et al. 2014). EMT is important in both development and tumour metastasis (Yang et al. 2008). Interestingly, adherens junctions

are involved in binding epithelial cells together (Yap et al. 1997) and the loss of these structures can cause transcriptional reprogramming. The adherens junction GO term was a top ten enriched cellular component GO term generated using UNR-associated proteins from U2OS plus arsenite (Table 4.19D) and UNR was shown to localize to areas of cell-cell interaction under certain conditions (Figures 4.22, 4.23).

Table 5.9: Seven putative UNR-interacting transcripts in arsenite stressed U2OS samples by two factor DESeq2 analysis

Genes	
ALCAM	*PLEKHA5*
BCAT1	*PRR11*
EDIL3	*ZNF608*
PIK3R1	

5.5.3 Results comparing U2OS (all samples) by immunoprecipitating antibody

Having seen few DESeq2-derived putative UNR-interacting RNAs from U2OS minus or plus arsenite, it was decided to merge the two and carry out DESeq2 analysis on all five pairs of U2OS samples.

As with the U2OS plus arsenite samples alone, one factor DESeq2 analysis failed to propose any putative UNR-interacting RNAs. Two factor analysis generated 30 significant hits, of which 12 were higher in the UNR samples. Of the 12, ten were protein coding (Table 5.10). Whilst none of the ten genes were particularly significant (all having BH-adjusted p-values between 0.04 and 0.05), two shared the lowest p-value (by the number of significant figures provided in the DESeq2 output); *PRR11* (see section 5.5.2) and *EFHC1*. EFHC1 has been linked to juvenile myoclonic epilepsy and is believed to be a regulator of mitosis and of neuronal migration during brain development (de Nijs et al. 2009). UNR has already been shown to be directly involved in neuronal migration in the developing brain (Kobayashi et al. 2013).

Another interesting hit was *ADAM19* which is involved in the pathogenesis of Alzheimer's disease (Tanabe et al. 2007). UNR was previously shown to be associated with the related *ADAM10* transcript which is also implicated in Alzheimer's (Renner 2014; Kim et al. 2009).

Table 5.10: Ten putative UNR-interacting transcripts in U2OS (all samples) by two factor DESeq2 analysis

Genes	
ACBD7	*FAM171B*
ADAM19	*PDGFA*
ARL10	*PRR11*
EFHC1	*RSRP1*
EGR1	*SERPINE1*

5.5.4 Summary of the DESeq2 analyses using U2OS samples

DESeq2 detected fewer putative UNR-interacting transcripts from the U2OS data than it did from the HeLa data. A possible reason for this is discussed below (see section 5.7.4). It was noted that the putative UNR-interacting *CXCL8* transcript from the U2OS minus arsenite samples was not suggested as a UNR interactor when DESeq2 carried out a two factor analysis of all U2OS samples. Also, of the seven putative UNR-interacting transcripts from the U2OS plus arsenite samples (Table 5.9), only one transcript, *PRR1* (see section 5.5.2), was a hit when all U2OS samples were analysed (Table 5.10). Interestingly, *PRR11* was also the only UNR-interactor suggested using U2OS cells that was also suggested as a hit using HeLa (HeLa plus arsenite).

Some evidence was found that strengthened suggestions in the literature linking UNR to Alzheimer's disease and metastasis. It should be noted that all cell types used in this experiment were immortalised cancer cell lines. Suggested links to cancer may say something about the function of UNR in non-cancer cells but it may be that the role of UNR in cancer cells is altered by the cellular reprogramming caused by the development of the cancer phenotype.

5.6 DESeq2 results for SaOS-2 samples

It was decided not to consider the SaOS-2 minus arsenite samples directly as there was only one SaOS-2 minus arsenite pair remaining after the removal of two IgG outliers and their paired UNR samples (Tables 5.2A, 5.2B). It was felt that the complete loss of biological replicates would make the risk of accepting false positives too high. This was, in part, due to the shape of the plot in Figure 5.7B.

5.6.1 Results comparing SaOS-2 plus arsenite samples by immunoprecipitating antibody

Two pairs of SaOS-2 plus arsenite samples remained after the removal of a single outlier and its paired sample (Tables 5.2A, 5.2B). One factor DESeq2 analysis of these samples suggested that 74 genes were significant at the BH-adjusted p-value cut off of 0.05 and that 10 of those were both protein coding and higher in UNR. The top hit was *CCNL2* with a BH-adjusted p-value of 0.0111.

The 10 genes did not yield any significantly over-represented GO terms. The genes are tabulated in Table 5.11.

Table 5.11: Ten putative UNR-interacting transcripts in arsenite stressed SaOS-2 by one factor DESeq2 analysis

Genes	
RBM10	*CEP95*
SRRT	*BRD8*
HGS	*PLCB4*
CCNL2	*KIAA0907*
MAPK8IP3	*SNRNP70*

As there were no significant GO terms, it was decided to carry out two factor analysis on these samples.

Two factor analysis generated 291 genes with BH-adjusted p-values less than 0.05, of which 121 were higher in the UNR samples. Of the 121 genes, 117 were protein coding. Feeding the 117 protein coding genes that were higher in the UNR samples into the GO tool yielded 26 significant biological process GO terms, 22 significant molecular function GO terms and 25 significant cellular component GO terms. The top ten GO terms for each class are provided in Table 5.12.

Table 5.12: Over-represented GO terms generated using putative UNR-interacting protein coding genes designated as significant at a BH-adjusted p-value cut off of 0.05 using two factor DESeq2 analysis of the SaOS-2 plus arsenite samples. (A) = biological process, (B) = molecular function, (C) = cellular component

A

GO biological process complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
mRNA metabolic process (GO:0016071)	613	21	3.42	6.14	2.55E-07
RNA splicing (GO:0008380)	373	17	2.08	8.17	3.17E-07
mRNA processing (GO:0006397)	436	18	2.43	7.4	3.96E-07
RNA processing (GO:0006396)	849	24	4.74	5.07	4.25E-07
macromolecule metabolic process (GO:0043170)	7359	73	41.05	1.78	1.37E-05
nucleic acid metabolic process (GO:0090304)	3942	50	21.99	2.27	1.82E-05
cellular macromolecule metabolic process (GO:0044260)	6693	68	37.34	1.82	3.89E-05
nucleobase-containing compound metabolic process (GO:0006139)	4484	53	25.02	2.12	5.52E-05
heterocycle metabolic process (GO:0046483)	4620	53	25.77	2.06	1.61E-04
cellular aromatic compound metabolic process (GO:0006725)	4669	53	26.05	2.03	2.34E-04

B

GO molecular function complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
RNA binding (GO:0003723)	1617	39	9.02	4.32	4.00E-12
poly(A) RNA binding (GO:0044822)	1159	33	6.47	5.1	1.11E-11
heterocyclic compound binding (GO:1901363)	5910	70	32.97	2.12	2.45E-09
nucleotide binding (GO:0000166)	2398	43	13.38	3.21	2.61E-09
nucleoside phosphate binding (GO:1901265)	2399	43	13.38	3.21	2.65E-09
organic cyclic compound binding (GO:0097159)	5991	70	33.42	2.09	4.96E-09
small molecule binding (GO:0036094)	2576	43	14.37	2.99	2.85E-08
protein binding (GO:0005515)	10751	95	59.98	1.58	4.25E-08
binding (GO:0005488)	14268	109	79.6	1.37	1.32E-07
nucleic acid binding (GO:0003676)	4039	53	22.53	2.35	3.62E-07

C

GO cellular component complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
nuclear lumen (GO:0031981)	3496	58	19.5	2.97	3.01E-13
nuclear part (GO:0044428)	3871	61	21.6	2.82	3.25E-13
nucleoplasm (GO:0005654)	2924	53	16.31	3.25	3.39E-13
nucleus (GO:0005634)	6893	80	38.46	2.08	5.62E-12
intracellular organelle lumen (GO:0070013)	4217	61	23.53	2.59	1.97E-11
organelle lumen (GO:0043233)	4289	61	23.93	2.55	4.38E-11
membrane-enclosed lumen (GO:0031974)	4344	61	24.23	2.52	7.96E-11
intracellular organelle (GO:0043229)	11885	103	66.3	1.55	2.26E-10
intracellular (GO:0005622)	14051	111	78.39	1.42	2.80E-10
intracellular organelle part (GO:0044446)	8056	84	44.94	1.87	3.20E-10

5.6.2 Results comparing all SaOS-2 UNR/IgG pairs by immunoprecipitating antibody

The SaOS-2 minus arsenite samples were then merged with the SaOS-2 plus arsenite samples to increase the number of overall samples and DESeq2 was then used to carry out another round of one factor analysis. This found 585 genes to have a BH-adjusted p-value less than 0.05, of which 265 were higher in the UNR samples. Of these 265 genes, 256 were protein coding. A log/log plot of the significant protein coding genes is given below (Figure 5.11).

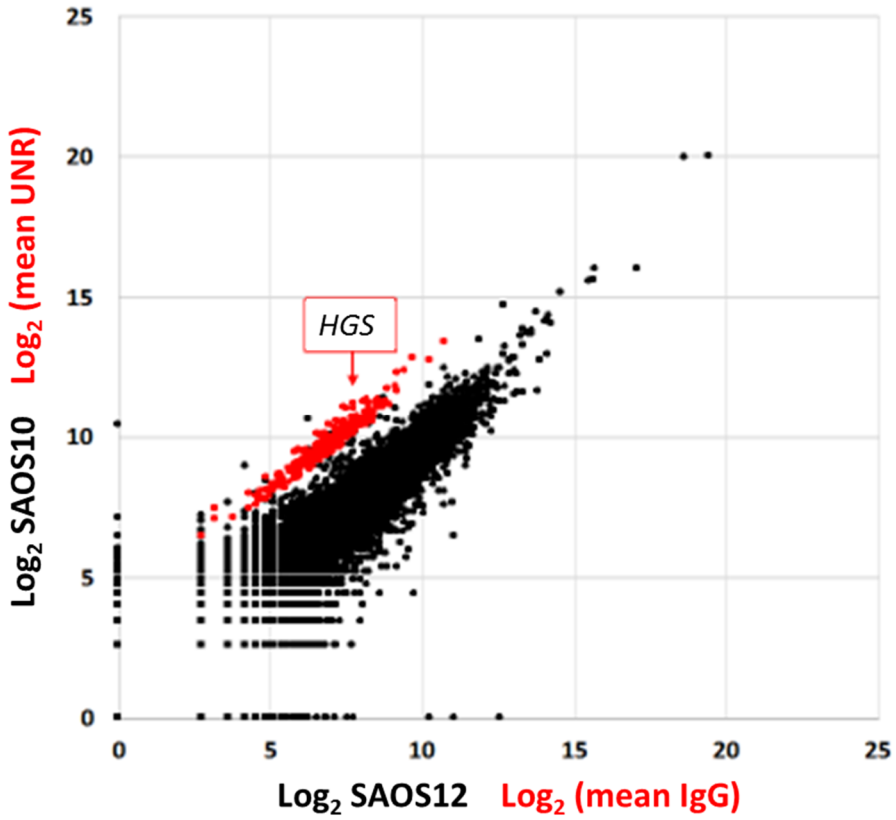


Figure 5.11: Log/log plot of one factor DESeq2 hits (red) and variability between two representative UNR pulldown samples (black) for all SaOS-2 samples. The name and location within the graph of the top hit is highlighted.

The suggested hits for all SaOS-2 formed a group that was separate from the main group of values given for the remaining SaOS-2 plus arsenite UNR samples, although there are some values that are more extreme in the background data (Figure 5.11). It was expected that there would be a clearer signal in SaOS-2 than in HeLa based upon the PCA analysis that showed the SaOS-2 IgG and UNR samples were clearly separated by PC2 (Figures 5.5A, 5.4A). It was decided to use the UNR plus arsenite samples to estimate the background variability in the data as UNR plus arsenite samples had been displayed for the HeLa samples.

As this analysis generated a large number of putative hits, it was considered interesting to consider the putative significant genes that had higher read counts in the IgG samples. If they were truly random, it would be expected that a lower proportion would be protein coding as UNR is a protein known to be involved in translational regulation. It was shown that, whereas 96% of the significant genes that had read counts higher in the UNR samples were protein

coding, only 74% of those higher in IgG were protein coding (238 out of 320 genes were protein coding). This finding was reassuring.

It was then decided to carry out GO term analysis on the protein coding genes that were higher in the UNR samples. This showed that there were 41 significantly over-represented biological process GO terms. Likewise, there were 31 significantly overrepresented molecular function GO terms and 40 significantly overrepresented cellular component GO terms. The top ten for each of these classes is presented in Table 5.13.

Table 5.13: Over-represented GO terms generated using putative UNR-interacting protein coding genes designated as significant at a BH-adjusted p-value cut off of 0.05 using one factor DESeq2 analysis on all SaOS-2 samples. (A) = biological process, (B) = molecular function, (C) = cellular component

A

GO biological process complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
nucleic acid metabolic process (GO:0090304)	3942	100	48.12	2.08	3.08E-10
cellular macromolecule metabolic process (GO:0044260)	6693	140	81.7	1.71	3.81E-10
nucleobase-containing compound metabolic process (GO:0006139)	4484	108	54.74	1.97	4.89E-10
RNA processing (GO:0006396)	849	41	10.36	3.96	6.21E-10
heterocycle metabolic process (GO:0046483)	4620	109	56.4	1.93	1.40E-09
cellular aromatic compound metabolic process (GO:0006725)	4669	109	56.99	1.91	2.88E-09
macromolecule metabolic process (GO:0043170)	7359	146	89.83	1.63	5.52E-09
organic cyclic compound metabolic process (GO:1901360)	4856	111	59.28	1.87	6.16E-09
cellular nitrogen compound metabolic process (GO:0034641)	5090	114	62.13	1.83	9.43E-09
cellular metabolic process (GO:0044237)	8525	160	104.06	1.54	1.24E-08

B

GO molecular function complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
protein binding (GO:0005515)	10751	207	131.23	1.58	1.80E-19
binding (GO:0005488)	14268	237	174.17	1.36	6.47E-18
RNA binding (GO:0003723)	1617	59	19.74	2.99	6.52E-11
poly(A) RNA binding (GO:0044822)	1159	48	14.15	3.39	3.12E-10
molecular_function (GO:0003674)	17018	247	207.73	1.19	4.03E-10
nucleotide binding (GO:0000166)	2398	72	29.27	2.46	7.38E-10
nucleoside phosphate binding (GO:1901265)	2399	72	29.28	2.46	7.53E-10
small molecule binding (GO:0036094)	2576	74	31.44	2.35	2.72E-09
organic cyclic compound binding (GO:0097159)	5991	127	73.13	1.74	2.75E-09
heterocyclic compound binding (GO:1901363)	5910	125	72.14	1.73	6.01E-09

C

GO cellular component complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
intracellular organelle part (GO:0044446)	8056	180	98.34	1.83	6.18E-22
organelle part (GO:0044422)	8240	181	100.58	1.8	3.31E-21
nuclear part (GO:0044428)	3871	118	47.25	2.5	6.79E-21
nuclear lumen (GO:0031981)	3496	111	42.67	2.6	1.44E-20
nucleoplasm (GO:0005654)	2924	99	35.69	2.77	1.48E-19
intracellular organelle lumen (GO:0070013)	4217	121	51.48	2.35	2.39E-19
organelle lumen (GO:0043233)	4289	121	52.35	2.31	1.05E-18
membrane-enclosed lumen (GO:0031974)	4344	121	53.03	2.28	3.15E-18
intracellular part (GO:0044424)	13672	232	166.89	1.39	4.02E-18
intracellular organelle (GO:0043229)	11885	215	145.08	1.48	1.10E-17

5.6.3 Summary of the DESeq2 analyses using SaOS-2 samples

It was decided not to consider the SaOS-2 minus arsenite samples as there was only one pair remaining after the removal of outliers and their pairs (Table 5.4). DESeq2 suggested that there were a number of putative UNR-interacting transcripts in the plus arsenite SaOS-2 samples, especially in the two factor analysis. Two factor analysis was not required for the combined SaOS-2 plus and minus arsenite samples due to a large number of putative hits from the one factor analysis. It was interesting that “nucleobase-containing compound metabolic process” was a top enriched biological process GO term for the SaOS-2 samples (plus arsenite or all samples), as for the HeLa samples. The most enriched biological process GO terms were largely related to binding to or modulating nucleic acids. This suggests that UNR could function as a modulator of RNAs both directly, as is already known, and indirectly by modulating the expression of other RNA-modulators. Further support is given to this hypothesis by the finding that the most enriched molecular function GO terms included “RNA binding” and “poly(A) RNA binding”. In terms of enriched cellular component GO terms, some of the top hits were related

to the nucleus. It is worth considering that mature transcripts do not tend to be found in the nucleus. This supports a role for UNR in modulating nuclear RNA processes such as splicing from the cytoplasm. Indeed, splicing was a top biological process GO term for the plus arsenite SaOS-2 samples. It is currently unknown why these findings were observed. Nevertheless, it is considered possible that UNR may have evolved to interact with and stabilise transcripts that are involved in splicing when a cell is stressed, and that it can then launch a translational response to the removal of the stress to help the cell recover by reprogramming its proteome following removal of the stress. In some ways, the increase in UNR observed during mitosis, coupled with it reprograming the proteome (e.g. by increasing translation of CDK11 through an IRES in its transcript (Schepens et al. 2007)), can be thought of as UNR in part mediating recovery from the stress of mitosis.

A direct comparison was made between the 111 UNR-interacting transcripts suggested in HeLa to the 256 suggested inn SaOS-2. Only three transcripts were common to both lists – *C1orf35*, *CLASRP* and *TADA2A*.

5.7 Cell type independent DESeq2 analysis

Having compared the samples by cell type and arsenite treatment, it was decided to look for additional hits by merging more samples together. There were three ways in which this was to be done:

1) All non-arsenite treated pairs irrespective of cell type
2) All arsenite treated pairs irrespective of cell type
3) All samples irrespective of both cell type and arsenite treatment.

It was clear by doing this that the number of repeats would go up. On the other hand, the PCA analysis showed that most of the variation in the data was related to cell type. For this reason, it was reasoned that information may come out of both one and two factor analyses. As discussed previously, the one factor analysis would be able to locate a subset of true hits that did not bind to IgG at any appreciable level (Figure 5.8A) or that were expressed at similar levels in all cell types (Figure 5.8B) but would not be so able to locate true hits where expression varied greatly between cell types (Figure 5.8C).

Two factor analysis, was considered more suited to detecting the differences in Figure 5.8C and was chosen over one factor analysis for consideration of the cell type independent samples.

5.7.1 DESeq2 results for all non-arsenite treated UNR/IgG sample pairs

Two factor DESeq2 analysis on all minus arsenite samples showed that only three genes had a BH-adjusted p-value less than 0.05, of which all three were putative UNR interactors. Two of these three genes were protein coding – *CCNL2* (adjusted p-value = 0.0076) and *LIMD1* (adjusted p-value = 0.0326). *CCNL2* has been mentioned previously as the top UNR hit from DESeq2 analysis of the SaOS-2 plus arsenite samples. As there were only two protein coding hits, GO term analysis was not attempted.

5.7.2 DESeq2 results for all arsenite treated UNR/IgG sample pairs

Two factor DESeq2 analysis of all arsenite treated samples following the removal of outliers and their paired IgG/UNR samples suggested that 369 genes were significant at the BH-adjusted $p<0.05$ significance level. 157 of these were higher in the UNR samples, of which 149 were protein coding. *PABPC1* was the top hit with a p-value of 3.11×10^{-11}. The 149 genes were fed into the GO tool which yielded a total of 67 significantly enriched biological process GO terms. By the hierarchical lay out option in the tool, it was seen that there were a lot of related terms, largely related with the regulation of transcription (e.g. direct positive or negative regulation of transcription or indirectly via processes such as histone/chromatin modification, data not shown). In total there were 67 significant biological GO terms. There were 14 significant

molecular function GO terms and 29 cellular component GO terms. The top ten most significant GO terms from each class, by Bonferroni-adjusted p-value, are tabulated below (Table 5.14).

Table 5.14: Over-represented GO terms generated using putative UNR-interacting protein coding genes designated as significant at a BH-adjusted p-value cut off of 0.05 using two factor DESeq2 analysis on all arsenite treated samples. (A) = biological process, (B) = molecular function, (C) = cellular component

A

GO biological process complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
chromosome organization (GO:0051276)	984	34	6.99	4.86	1.13E-10
nucleic acid metabolic process (GO:0090304)	3942	62	28.01	2.21	9.04E-07
nucleobase-containing compound metabolic process (GO:0006139)	4484	67	31.86	2.1	9.19E-07
heterocycle metabolic process (GO:0046483)	4620	68	32.82	2.07	1.18E-06
cellular macromolecule metabolic process (GO:0044260)	6693	85	47.55	1.79	1.87E-06
cellular aromatic compound metabolic process (GO:0006725)	4669	68	33.17	2.05	1.91E-06
negative regulation of gene expression (GO:0010629)	1412	34	10.03	3.39	2.21E-06
chromatin modification (GO:0016568)	289	16	2.05	7.79	2.96E-06
cellular nitrogen compound metabolic process (GO:0034641)	5090	71	36.16	1.96	4.06E-06
macromolecule metabolic process (GO:0043170)	7359	89	52.28	1.7	6.30E-06

B

GO molecular function complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
binding (GO:0005488)	14268	133	101.37	1.31	2.83E-06
protein binding (GO:0005515)	10751	112	76.38	1.47	4.87E-06
nucleic acid binding (GO:0003676)	4039	57	28.7	1.99	1.48E-04
heterocyclic compound binding (GO:1901363)	5910	73	41.99	1.74	1.61E-04
organic cyclic compound binding (GO:0097159)	5991	73	42.56	1.72	2.93E-04
RNA binding (GO:0003723)	1617	31	11.49	2.7	8.95E-04
nucleotide binding (GO:0000166)	2398	39	17.04	2.29	1.38E-03
nucleoside phosphate binding (GO:1901265)	2399	39	17.04	2.29	1.39E-03
chromatin binding (GO:0003682)	450	15	3.2	4.69	2.37E-03
small molecule binding (GO:0036094)	2576	40	18.3	2.19	3.08E-03

C

GO cellular component complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
nucleoplasm (GO:0005654)	2924	65	20.77	3.13	1.76E-15
nuclear lumen (GO:0031981)	3496	71	24.84	2.86	2.40E-15
nuclear part (GO:0044428)	3871	73	27.5	2.65	3.66E-14
organelle lumen (GO:0043233)	4289	75	30.47	2.46	6.90E-13
intracellular organelle lumen (GO:0070013)	4217	74	29.96	2.47	1.06E-12
membrane-enclosed lumen (GO:0031974)	4344	75	30.86	2.43	1.41E-12
intracellular organelle part (GO:0044446)	8056	105	57.24	1.83	2.86E-12
nucleus (GO:0005634)	6893	96	48.97	1.96	4.19E-12
intracellular part (GO:0044424)	13672	138	97.14	1.42	4.20E-12
organelle part (GO:0044422)	8240	105	58.54	1.79	1.65E-11

Chromosome organization, chromosome modification and negative regulation of gene expression were among the top biological process GO terms. This is interesting in light of the hypothesis put forward in section 5.6.3 that stated UNR could be involved in stabilising transcripts during stress that are later involved in recovery from stress. As well as a translational control of proteins involved in splicing and other RNA modifications, UNR may also modulate the transcriptional output in response to the removal of a stress. This would not be the first observation linking UNR to the regulation of transcription (e.g. UNR was shown to interact directly with the histone methyl transferase transcriptional regulator ALL-1 [Leshkowitz et al. 1996]).

The top ten most enriched molecular function GO terms, whilst including very general GO terms like binding, contained chromatin binding. There were also 57 hits for nucleic acid binding but only 31 for RNA binding, further implying that UNR may be interacting with transcripts encoding transcription factors.

The most enriched cellular component GO terms included ‘nucleus’, with 96 associated genes. As well as ‘nucleus’, the top three hits were also nucleus-related.

5.7.3 DESeq2 results for all UNR/IgG sample pairs

Finally, it was decided to merge all samples together, irrespective of both their cell type and arsenite treatment state. This merger gave the highest possible number of repeats for paired two factor analysis. On the other hand, it also maximised the effect of the two major confounding variables – cell type and arsenite treatment. To discuss that effect directly, the identifier table inputted into DESeq2 is provided in Table 5.15.

Table 5.15: Identifiers for all samples for two factor DESeq2 analysis

Sample	RNA-Seq ID	condition	rep
HeLa2	238128_001	I	H1
HeLa3	238128_004	I	H2
HeLa5	238128_008	U	H1
HeLa6	238128_006	U	H2
HeLa7	238129_007	I	H0
HeLa8	238128_010	I	H1
HeLa9	238129_012	I	H2
HeLa10	238129_013	U	H0
HeLa11	238128_011	U	H1
HeLa12	238128_014	U	H2
SaOS2	238129_003	I	S1
SaOS5	238129_010	U	S1
SaOS7	238129_020	I	S0
SaOS9	238129_002	I	S2
SaOS10	238129_004	U	S0
SaOS12	238129_006	U	S2
U2OS1	238128_015	I	U0
U2OS2	238128_016	I	U1
U2OS3	238128_018	I	U2
U2OS4	238128_019	U	U0
U2OS5	238128_003	U	U1
U2OS6	238128_009	U	U2
U2OS7	238128_020	I	U0
U2OS9	238128_022	I	U2
U2OS10	238129_023	U	U0
U2OS12	238129_027	U	U2

As with all previous analyses, the first (or only) factor considered by DESeq2 was provided in the "condition" column (e.g. whether the samples were IgG [I] or UNR [U]). Also, as with analyses

on all samples from individual cell types, the plus and minus arsenite samples made on the same day were given the same “rep” value where “rep” was the second factor used in two factor analysis. DESeq2 was therefore programmed to consider every day upon which samples were made as a factor. For SaOS-2, all the samples were produced on separate days as the pairs lost to outliers in the minus arsenite sample were complementary to the pair lost to an outlier in the plus arsenite samples. On the other hand, there were four “rep” identifiers with four members (H1, H2, U0 and U2). There were therefore variable degrees of mixture of plus and minus arsenite samples – both overall and within specific “rep” values. On top of this, the majority of the variability had been shown by PCA analysis to have been related to cell type differences and merging all samples together mixes up the cell types. That was the reason for using two factor analysis; looking for differences between UNR/IgG differences between samples made on the same day. Unlike the plus or minus arsenite samples, however, the plus and minus arsenite data has the additional confounding effect of arsenite treatment. In summary, it was not expected to find many large fold change differences but it was hoped that a number of smaller fold changes would nevertheless prove significant due to the larger number of repeats.

The two factor DESeq2 analysis suggested that 829 genes were significant, of which 454 were higher in the UNR samples, of which 417 were protein coding. GO term analysis yielded 61 significantly enriched biological process GO terms, 37 significantly enriched molecular function GO terms and 35 significantly enriched cellular component GO terms. The top ten from each class is tabulated in Table 5.16. Other observations from less significant biological process GO terms implied a role for proteins translated from UNR-interacting transcripts in RNA metabolism, mitosis and gene expression (data not shown). Likewise, there was evidence for a role for proteins encoded by UNR-interacting transcripts in molecular functions such as S-adenosylmethionine-dependent methyltransferase activity, chromatin binding and DNA binding (data not shown). In terms of cellular component GO terms, UNR was implicated in binding to transcripts that encode centrosomal, nucleolar and spindle proteins (data not shown).

Table 5.16: Over-represented GO terms generated using putative UNR-interacting protein coding genes designated as significant at a BH-adjusted p-value cut off of 0.05 using two factor DESeq2 analysis on all samples. (A) = biological process, (B) = molecular function, (C) = cellular component

A

GO biological process complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
nucleobase-containing compound metabolic process (GO:0006139)	4484	160	89.16	1.79	1.97E-11
nucleic acid metabolic process (GO:0090304)	3942	146	78.38	1.86	3.42E-11
heterocycle metabolic process (GO:0046483)	4620	161	91.86	1.75	1.29E-10
cellular aromatic compound metabolic process (GO:0006725)	4669	161	92.84	1.73	3.30E-10
organic cyclic compound metabolic process (GO:1901360)	4856	163	96.56	1.69	2.21E-09
cellular nitrogen compound metabolic process (GO:0034641)	5090	168	101.21	1.66	3.18E-09
cellular macromolecule metabolic process (GO:0044260)	6693	204	133.08	1.53	3.35E-09
chromosome organization (GO:0051276)	984	57	19.57	2.91	6.52E-09
organelle organization (GO:0006996)	3133	117	62.3	1.88	3.94E-08
chromatin modification (GO:0016568)	289	28	5.75	4.87	1.07E-07

B

GO molecular function complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
binding (GO:0005488)	14268	358	283.7	1.26	1.16E-13
protein binding (GO:0005515)	10751	295	213.77	1.38	1.16E-12
organic cyclic compound binding (GO:0097159)	5991	193	119.12	1.62	3.59E-11
heterocyclic compound binding (GO:1901363)	5910	191	117.51	1.63	4.11E-11
nucleic acid binding (GO:0003676)	4039	145	80.31	1.81	1.88E-10
RNA binding (GO:0003723)	1617	77	32.15	2.39	2.51E-09
poly(A) RNA binding (GO:0044822)	1159	57	23.05	2.47	1.08E-06
nucleotide binding (GO:0000166)	2398	92	47.68	1.93	1.33E-06
nucleoside phosphate binding (GO:1901265)	2399	92	47.7	1.93	1.36E-06
small molecule binding (GO:0036094)	2576	96	51.22	1.87	2.29E-06

C

GO cellular component complete	Reference	Observed	Expected	Fold increase	Adjusted p-value
nuclear part (GO:0044428)	3871	170	76.97	2.21	4.51E-23
nuclear lumen (GO:0031981)	3496	158	69.51	2.27	3.89E-22
nucleoplasm (GO:0005654)	2924	140	58.14	2.41	3.94E-21
intracellular part (GO:0044424)	13672	363	271.85	1.34	4.03E-21
intracellular (GO:0005622)	14051	366	279.39	1.31	1.19E-19
nucleus (GO:0005634)	6893	234	137.06	1.71	2.11E-19
intracellular organelle lumen (GO:0070013)	4217	163	83.85	1.94	6.02E-16
intracellular organelle (GO:0043229)	11885	323	236.32	1.37	6.53E-16
organelle lumen (GO:0043233)	4289	164	85.28	1.92	1.28E-15
membrane-enclosed lumen (GO:0031974)	4344	164	86.37	1.9	4.52E-15

As with the plus arsenite samples, chromatin organization and chromatin modification were among the top ten enriched biological process GO terms. In both cases, the number of significant genes annotated to the GO terms increased (57 versus 34 and 28 versus 16, respectively). Nevertheless, the observed/expected fold increases decreased in both cases. The p-values were still strong, and actually decreased in the case of chromatin modification. Beyond the chromatin-related GO terms, most of the other top ten biological process GO terms were also similar to those for the plus arsenite samples alone. Chromatin binding was not on the top ten enriched as it was for the plus arsenite samples. The strongest p-values were for binding and protein binding, with 358 and 295 genes annotated, respectively. There were 145 genes annotated to nucleic acid binding and only 77 to RNA binding, again implying that at least some of the transcripts to which UNR binds may encode proteins that can bind to DNA. As with the plus arsenite samples, the top three enriched cellular component GO terms by p-value were nuclear related and nucleus, to which 234 genes were annotated, was in sixth place with a p-value of 2.11×10^{-19}.

5.7.4 Review of DESeq2 results for cell independent samples

The first main observation was that there were few hits for the minus arsenite samples (section 5.7.1), both in absolute terms and relative to the plus arsenite samples (section 5.7.2). One possible explanation for this finding is that UNR binds to fewer transcripts under non-stress conditions (possibly due to there being less UNR protein in the cell or to UNR being sequestered away from the transcripts to which it binds following arsenite stress). Stresses, such as oxidative stress with arsenite or the stress associated with mitosis could result in UNR levels increasing and/or UNR binding to different groups of transcripts.

In terms of mitosis, it has been reported that UNR represses its own translation until hnRNP C1/C2 proteins are released from the nucleus following the breakdown of the nuclear envelope. These then replace UNR on its transcript and drive its translation (Schepens et al. 2007). This results in an increase in free UNR that can then bind to certain transcripts, among them cdk11 (Tinton et al. 2005). The problem with discovering the identity of mitosis-specific UNR-interactors was that the cells used in the experiments had been grown to 70-80% confluent. That means that relatively few would have been undergoing mitosis at the point of lysis and any differences may have been diluted by the majority of UNR-interactors coming from cells that were either in G_0 or some other phase of the cell cycle prior to nuclear envelope breakdown.

In terms of arsenite stress, it has been shown that overexpressed UNR goes to stress granules in HeLa cells following arsenite stress (White & Lloyd 2011). Furthermore, work in the Anderson lab has confirmed that UNR is found in the numerous stress granules that form following arsenite stress in TP53 compromised cells (HeLa and SaOS-2). Many fewer stress granules were found in wild type TP53 cells (U2OS). It was hypothesised that UNR was part of active translation complexes that are taken to stress granules upon arsenite stress (Ray, Ó Catnaigh and Anderson, manuscript in preparation). The larger number of UNR-associated HeLa transcripts found in the plus arsenite samples relative to the minus arsenite samples could be related to the ease of pulling down multiple UNR-containing complexes if they were concentrated in stress granules. It was also noted that the *TP53*-null SaOS-2 plus arsenite samples had more hits than the wild type *TP53* U2OS plus arsenite samples, even though they had the same number of pairs (Table 5.4).

Hits detected following the merging of samples from all cell types were interesting in that they implied a role for UNR that may be cell type independent. Many of the significant GO terms in the plus arsenite were related to transcription and the cell cycle. It is fair to infer that, if a cell survives a particular stress, it would have to change its transcriptional output to revert from having a proteome that is geared towards surviving the stress to one that is more geared towards an unstressed state. These findings could suggest that UNR is involved in protecting a specific subset of transcripts following stress with arsenite. They are also in keeping with the hypothesis, as postulated in (Ray, Ó Catnaigh and Anderson, manuscript in preparation) that UNR is a general translation factor. There are three possible situations:

1) UNR is bound to active translation complexes that are all translocated to stress granules upon arsenite stress (possibly via UNRIP).
2) UNR is bound to active translation complexes, a subset of which are translocated to stress granules upon arsenite stress (possibly via UNRIP).
3) UNR is not bound to active translation complexes but recruits a subset of them for translocation to stress granules following arsenite stress.

The merging of all samples, independent of both arsenite treatment and cell type, provided a slightly different set of results. Some of the suggested UNR-interacting transcripts, for example, may be strong hits in the plus arsenite samples but not hits in the minus arsenite samples. Others may be due to hits that were genuinely higher in UNR across all samples. Looking at the top biological process GO term for the plus arsenite samples – "chromosome organization", it had an adjusted p-value of around 1.1×10^{-10}, 34 observed proteins and a fold enrichment of just under 5 (Table 5.14A). The same GO term was the eighth highest when putative UNR-interacting transcripts calculated using two factor DESeq2 analysis of all samples. The adjusted p-value did become less significant – raising to around 6.5×10^{-9} and the fold change fell to just under 3 (Table 5.16A). On the other hand, there were then 57 observed proteins. These observations imply that the actual situation, at least for this GO term, may be affected by both considerations. DESeq2 detected almost twice as many putative UNR-interacting transcripts that are annotated to that GO term when all samples were merged over when only the plus arsenite samples were merged. At the same time, the distribution of transcripts changed and the proportion annotated to that GO term fell. Comparing Tables 5.14 and 5.16 in their entireties, it seems likely that UNR has a role in binding to transcripts involved in processes such as gene expression and mitosis,

functions such as RNA binding and that tend to be located in the nucleus. The slight differences between the significant GO terms for all samples and the plus arsenite samples only implied that UNR may be more involved in binding to mitosis-related transcripts when unstressed. As an example, *CDK11A* (Tinton et al. 2005) had an adjusted p-value of 0.0249 when all samples were merged but the same gene had an adjusted p-value of 0.6589 in the plus arsenite samples.

5.8 General summary of DESeq2 results

A number of putative UNR-interacting, protein coding RNAs was discovered using RIP-Seq and post-experimental analysis with DESeq2. Among the RNAs that were previously shown to interact with UNR, there were not many calculated to be significant by DESeq2. Indeed, of the 7 mammalian RNAs mentioned in (Ray, Ó Catnaigh, & Anderson, 2015), only two were considered significant under any conditions. Parathyroid hormone receptor was not present in any of the datasets. Some of the data for the other transcripts is presented below (Table 5.17).

Table 5.17: DESeq2 adjusted p-values for 6 known UNR-interacting transcripts over a selection of conditions with significant BH-adjusted p-values in yellow

Analysis:	One factor		Two factor	
	HeLa +As	HeLa (all)	Plus As	All samples
Transcript	BH-adjusted p-values			
APAF1	0.9744	0.9571	0.9401	0.7268
CDK11	-	0.4355	0.6589	0.0249
FOS	0.9812	0.9985	0.4602	0.6168
GATA6	-	0.9171	0.9974	0.9580
PABP1	2.52E-12	2.36E-14	3.11E-11	1.08E-05
UNR	0.7865	0.8472	0.9808	0.9713

Given that many known UNR-associated transcripts were not found to be significant by DESeq2, it could be that these results, the published results, or both results are incorrect (Ioannidis 2005).

Alternatively, UNR may bind to many more transcripts than is currently known and the small scale observations previously made, whilst true, may be a small subset of the total array of transcripts to which UNR actually does bind. If that is true, then it may not be surprising that not all hits would be seen in a high throughput experiment such as was carried out for this work. On top of this, some of the known interactors were context dependent (e.g. pertaining to apoptosis, mitosis, development in mouse embryonic fibroblasts, etc.).

5.9 Future direction for study

It would be helpful to carry out a second round of RIPs followed by qPCR on some if not all of the DESeq2-discovered putative UNR-interacting transcripts with a view to validating them. Even in the absence of qPCR, it would be interesting to see if a repeat of these experiments yielded the same results. It would also be useful to check all putative UNR-interacting transcripts both for IRES sequences and known UNR binding motifs. Putative UNR-interacting transcripts may not be direct interacting transcripts but part of an overall RNP that contains UNR. It would therefore be useful to ascertain how many of the putative UNR-interacting transcripts are both real and direct interactors. Cdk11, for example, is known to possess an IRES through which UNR directs translation at a specific part of the cell cycle. It could be that UNR acts via IRES structures on a wide range of transcripts.

It was also noted that some non-coding RNAs were considered significant in some of the DESeq2 analyses. As a result of this project being geared towards discovering UNR-interacting proteins and transcripts, a body of information was left uninvestigated. As more information is discovered about non-coding RNAs and suitable bioinformatics tools become available to process them, it would be interesting to see if UNR has any role in modulating cellular functions via non-coding RNAs.

6 Identification of proteins with expression levels that are modulated by UNR

Having explored UNR-interacting proteins and transcripts, it was then decided to carry out *UNR* knockdown experiments (see sections 6.1 and 6.2) followed by mass spectrometry in an attempt to discover proteins and/or groups of proteins that have their expression level modulated by UNR. These experiments were based upon the assumption that, if a protein is differentially expressed when UNR is reduced to relatively very low levels, it is likely that the presence of UNR affects the expression of that protein.

6.1 RNA interference and gene knockdown

Gene knockdown is a technique that takes advantage of the phenomenon of RNA interference to greatly reduce the expression of targeted proteins by causing the degradation of their transcripts.

RNA interference is currently understood to have three pathways based upon which of three groups of small RNA molecule is involved. There are three such groups – PIWI-interacting RNA (piRNA), microRNA (miRNA) and small interfering RNA (siRNA) (reviewed by Wilson & Doudna 2013). In each case, the small RNA interacts with one of the four Argonaute proteins (in humans); AGO1, AGO2, AGO3 and AGO4 (reviewed in Ender & Meister 2010). The three groups of small RNA molecule differ in terms of their origin and the processing steps required before they can integrate into functional RNA-induced silencing complexes (RISC) (Kobayashi & Tomari 2016). Further consideration will be given to the miRNA and siRNA pathways. miRNAs are transcribed from endogenous genes as hairpin loop structure-containing primary miRNAs (Kobayashi & Tomari 2016). These then have their stems cleaved in the nucleus by the DROSHA-containing microprocessor complex, leaving a stem of around 25 imperfectly matched base pairs with a two nucleotide 3’ overhang and a small hairpin loop of around 10 nucleotides (Lee et al. 2003; Denli et al. 2004; Kobayashi & Tomari 2016). A group of primary miRNAs only have a one nucleotide 3’ overhang and require the addition of a 3’ uridine (Heo et al. 2012). Exportin 5-mediated transport to the cytoplasm then occurs where DICER cleaves off the hairpin loop, leaving an imperfectly base paired double stranded RNA molecule with 2 nucleotide 3’ overhangs (Wilson & Doudna 2013; Kobayashi & Tomari 2016). This interacts with an Argonaute protein as part of an miRISC complex, where one strand (the guide strand) will be retained to act as a substrate

recognition molecule for target mRNAs and the other (the passenger stand) is ejected. This can then lead to translational repression and/or transcript degradation via slicing and deadenylation (Jonas & Izaurralde 2015).

Unlike miRNA, siRNA can have an exogenous source such as being formed from the cleavage of viral dsRNA (Ender & Meister 2010). Endogenous sources of siRNA have also been reported, such as antisense transcript products of pseudogenes that can interact with sense transcripts of the original gene to generate a duplex (Tam et al. 2008; Ender & Meister 2010). In relation to this source of siRNA, it is interesting to consider the presence of *UNR*-like sequences in other regions of the human genome (see section 1.1.5). As with pre-miRNA, DICER processes pre-siRNA duplexes and generates duplexes with 3' overhangs (Carthew & Sontheimer 2009). Unlike miRNA, siRNA base pairing is perfectly complementary (Carthew & Sontheimer 2009). Both miRNA and siRNA duplexes can activate all four Argonaute proteins to generate functional miRNA/siRNA-RISC complexes following passenger strand ejection that can then function in gene silencing (Nakanishi 2016).

Exogenous siRNA duplexes can be transfected into transfectable cells to activate gene silencing (Elbashir et al. 2002; O'Keefe 2013). The experiments in this chapter were carried out using a commercial siRNA against *UNR* or a commercially available control siRNA that does not have a known target in the human transcriptome (see section 2.1.2). The si*UNR* was used to reduce the amount of *UNR* mRNA and, over time, reduce the cellular level of the UNR protein.

6.2 Experimental approach

In order to do this, two experiments were devised. In the first, unstressed HeLa cells were used and, in the second, unstressed and arsenite stressed U2OS cells were used. In both cases, cells were split into 10cm tissue culture plates and were allowed to grow to a confluency of 50-70%. A lipofectamine-based knockdown was then undertaken. Briefly, for each plate, 1 ml of Opti-MEM was mixed with 20 µl of Lipofectamine 2000 and left for 5 minutes at room temperature. Meanwhile, 2 µl of either si*UNR* or control siRNA were added to another 1 ml of Opti-MEM. The two solutions were then mixed by inversion and left for 20 minutes at room temperature. During this time, the cells were fed with 18 ml of fresh DMEM containing 10% FBS. Finally, the Opti-MEM mixture was added dropwise around the plates which were then returned to the incubator

for 42-48 hours. After this time, the cells were either treated with 1 mM sodium arsenite in fresh DMEM or a similar volume of sterile PBS in fresh DMEM for one hour prior to harvest. Cell harvesting was carried out using polysome lysis buffer without RNase OUT. A small amount of lysate (about 5%) was removed to run on a Western and the rest was frozen at -80°C prior to preparation for analysis by mass spectrometry.

A different protocol was used for processing these samples for mass spectrometry than those that had been used in Chapter 4. The protocol used was the standard 'Filter Aided Sample Preparation' (FASP) method used in the Life Sciences proteomics facility at the University of Warwick which was adapted from Wiśniewski et al. (2009).

The HeLa samples were run as 15 µl injections for 4 hours and the U2OS samples were run as 10 µl injections for 3 hours.

6.3 *UNR* knockdown in HeLa cells

6.3.1 Only two repeats were considered

A trial run was carried out using the new mass spectrometry preparation method, followed by two further repeats. Unfortunately, the proteomics department changed their FASP protocol slightly between the two sets. According to an explicit warning stated within the protocol, there is an essential acidification step that promotes the binding of peptides to the C18 column that is used to collect peptides prior to elution and analysis by mass spectrometry. An insufficient amount of acid had been stated in the earlier version of the protocol and this was increased twenty-fold in the updated version. When the two groups of samples were considered together by the proteomics department, they advised that there were significant differences between the trial run and the two subsequent repeats (data not shown). It was therefore decided to present only the data from the latter two repeats.

6.3.2 Western blot confirmation of successful knockdown

As stated previously, a proportion (approximately 1/20) of the cell lysates were retained for analysis by Western blot in order to confirm that the knockdown had been successful. In the case of the samples in question, the knockdowns had worked as expected (Figure 6.1)

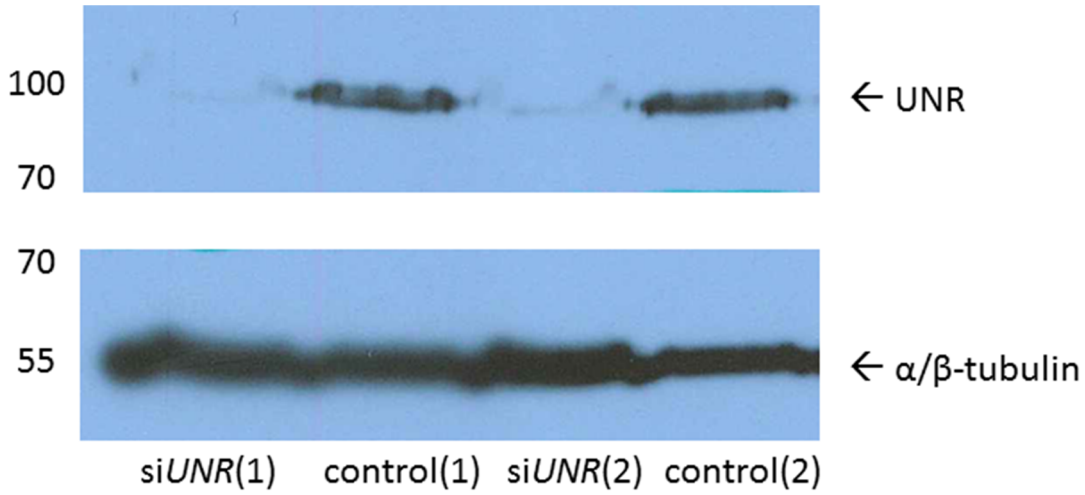


Figure 6.1: Western blot showing reduction in UNR levels following *UNR* knockdown. HeLa cells were treated with si*UNR* or a control siRNA (as stated) 48 hours prior to harvesting. The images show UNR (top panel) or α/β-tubulin (lower panel) and come from a single gel that was transferred to nitrocellulose membrane that was cut at a 70kDa marker so as to allow probing for both proteins.

6.3.3 Progenesis analysis of HeLa *UNR* knockdown data

It was previously stated that the use of the Progenesis software was preferred over Scaffold. As a result of that, it was decided to only provide Progenesis data for the HeLa samples. As before, files containing raw data from the mass spectrometer for each sample to be compared were loaded into Progenesis and processed with the aid of Mascot (see section 4.8). Automatic alignment showed that all samples were well aligned and there were 141066 features (i.e. putative peptides) in total, 2243 of which were removed when a maximum charge of 6 and maximum run time of 225 minutes was employed. Features with no MS/MS data or no protein ID were removed, as were peptides with Mascot scores less than 25 and those assigned to

keratins, pig proteins and bovine proteins. As proteins had not been immunoprecipitated in this experiment, immunoglobulins were not removed from the data.

An image was made of principal component 1 against principal component 2 from a PCA analysis using suggested proteins using all data (Figure 6.2A). A second PCA analysis was then carried out on the remaining data after all proteins with ANOVA p-values greater than or equal to 0.1 were removed (Figure 6.2B).

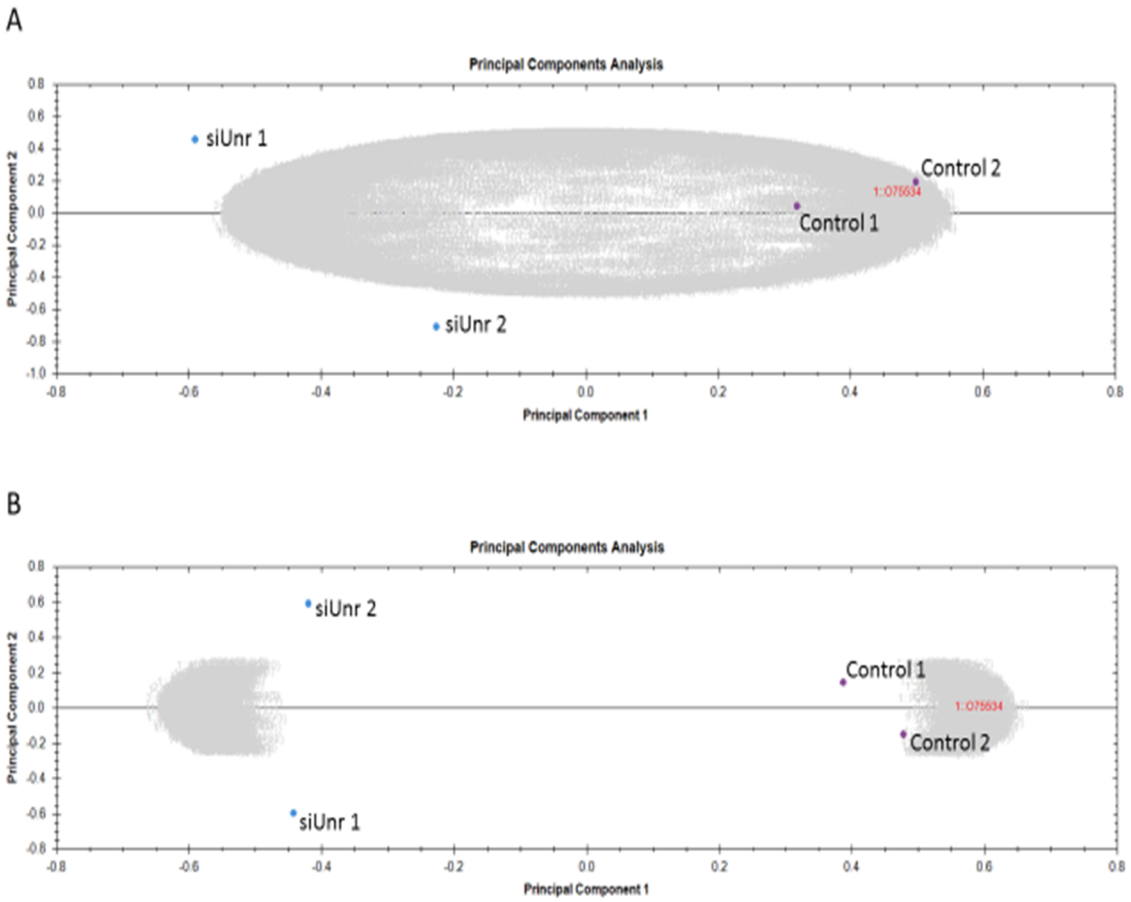


Figure 6.2: Progenesis-generated PCA plots for HeLa knockdown data. Proteins with all ANOVA p-values (A) or only those with ANOVA p-values under 0.1 (B) were included. See text for further information on how the data were processed. Numbers are associated with the days upon which the lysates were made (i.e. the repeats). Blue = si*UNR* samples (siUnr); purple = control siRNA samples (Control). In terms of variance accounted for – A: (PC1 = 50.96%, PC2 = 35.29%); B: (PC1 = 91.54%, PC2 = 4.83%). The grey writing in the background makes up the protein score plot (see section 4.9.2). The red writing refers to an automatically highlighted protein ID, selected to be UNR.

Examination of the PCA plots showed that PC1 accounted for over 50% of the total variability in both cases and clearly divided the samples by whether they were si*UNR* or control treated. The source of variation within PC2 was less obvious. Whereas PC2 accounted for over 35% of the total variability when all proteins were considered, it only accounted for around 3.6% when proteins with ANOVA p-values greater than or equal to 0.1 were excluded. There was an argument not to exclude proteins based on their ANOVA values alone because there were only two groups, thereby effectively reducing the ANOVA to an unpaired t-test. The proteins were removed nonetheless because:

a) Removing them greatly increased the amount of variability accounted for by PC1 (which was closely associated with the treatment – i.e. they were well separated by the line PC1 =0), and

b) Paired and unpaired t-tests ask different questions of groups of samples and it seemed useful to remove proteins that did not have any particular difference in their means. It was considered possible that using both tests could potentially remove some false positives at the cost of removing some weaker true positives.

Although it had already been shown that the knockdown had worked well (Figure 6.1), it was considered useful to confirm that the reduction in UNR was also apparent in the Progenesis analysis. A graphical confirmation of the knockdown showed that the peptides assigned to UNR all followed a similar distribution between the samples, corroborating the idea that they came from the same protein, and were higher in the control siRNA sample than in the si*UNR* samples (Figure 6.3).

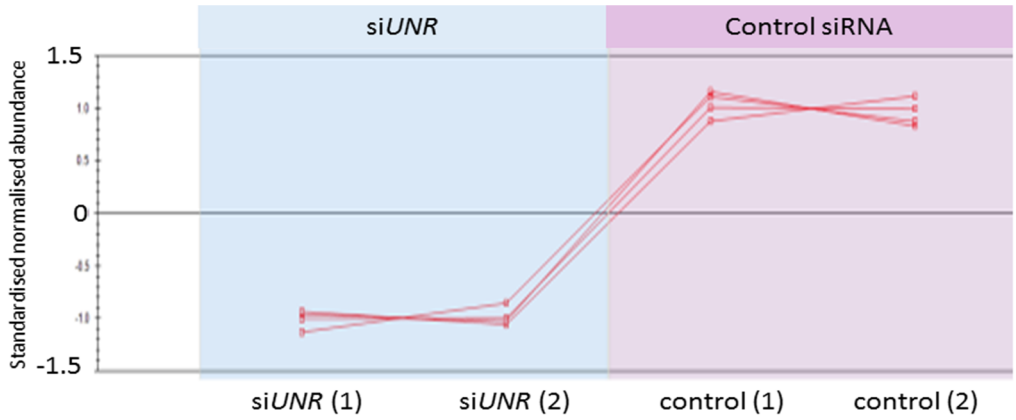


Figure 6.3: Standardised normalised abundances for peptides assigned to UNR from the HeLa knockdown samples. The repeats in the blue column are the si*UNR*-treated samples and the repeats in the purple column are the control siRNA-treated samples. This image was produced using Progenesis.

6.3.4 Detection of proteins that are differentially expressed following *UNR* knockdown in HeLa cells

The protein abundance data was then exported to Microsoft Excel and paired t-tests were carried out between the si*UNR* and corresponding control siRNA samples for each observed protein that passed the ANOVA p-value <0.1 significance level cut off. Ratios of the sum of the si*UNR*-treated sample abundances for each protein to the sum of the sample abundances for the control siRNA-treated sample abundances were also calculated for each observed protein. From those calculations, putative proteins with UNR-mediated expression were taken to be those with:

1) t-test p-value less than 0.05, higher in si*UNR* samples
2) t-test p-value less than 0.05, higher in control siRNA samples
3) si*UNR*/control siRNA ratio greater than 10*
4) si*UNR*/control siRNA ratio less than 0.1*

* if the ratio was infinite, only proteins with both si*UNR* values greater than zero were accepted and, if the ratio was zero, only proteins with both control siRNA values greater than zero were accepted.

It should be explicitly noted here that the t-test p-values were not adjusted using a multiple testing correction.

6.3.5 Discovery of proteins whose expression level changes on si*UNR* treatment in unstressed HeLa cells

The changes in protein levels were explored in turn for each of the four groups stated in the previous section.

1) t-test p-value less than 0.05, higher in si*UNR* samples

In total, there were 128 proteins that had p-values under 0.05 and higher abundances in the si*UNR*-treated samples. These may have included multiple versions of a given protein but any such duplicates were later removed prior to carrying out GO-term overrepresentation analysis on the putative hits (*v.i.*). The top ten hits that were higher in the si*UNR*-treated samples, by t-test p-value, are presented in Table 6.1.

Table 6.1: Top ten proteins with higher expression in si*UNR*-treated HeLa cells than in control siRNA-treated HeLa cells, by t-test p-value

Protein	p-value
Thymidylate synthase GN=TYMS	0.00093
Isoform 2 of 60S ribosomal protein L11 GN=RPL11	0.00168
Macrophage-capping protein GN=CAPG	0.00210
Translation machinery-associated protein 7 GN=TMA7	0.00224
TRPM8 channel-associated factor 1 GN=TCAF1	0.00235
Adenosylhomocysteinase GN=AHCY	0.00246
Tubulin gamma-2 chain GN=TUBG2	0.00254
Transient receptor potential cation channel subfamily V member 2 GN=TRPV2	0.00307
Isoform 2 of Protein arginine N-methyltransferase 5 GN=PRMT5	0.00339
Nucleosome assembly protein 1-like 1 (Fragment) GN=NAP1L1	0.00371

The top hit, thymidylate synthase (TYMS), is interesting (Table 6.1). This oncogene encodes a protein that is essential for the *de novo* production of deoxythymidine-5'-monophoshate and, through that, for DNA synthesis. Its upregulation is associated with a poor prognosis in a number of cancers (Rahman et al. 2004; Lenz et al. 1996; Popat et al. 2004). It is possible that the protein is present in higher amounts in U2OS cells due to their malignant phenotype. Nevertheless, seeing a rise in TYMS levels following si*UNR* treatment suggests that UNR could possibly function as a tumour suppressor by keeping the levels of TYMS low.

The second top hit was the ribosomal protein, RPL11 (Table 6.1). That protein was shown to be associated with Diamond Blackfan anaemia (Quarello et al. 2010). That finding links into the UNR-related work being carried out in the Von Lindern lab (Horos et al. 2012). The protein also regulates TP53 activity through HDM2 (Lohrum et al. 2003). Three other genes from Table 6.1 are TUBG2, TCAF1 and TRPV2 (Table 6.1). TUBG2 encodes a subunit of γ-tubulin. Noting a link between UNR levels and γ-tubulin at this point was interesting in that the lab had previously used γ-tubulin as a loading control for Western blots. It was noted that confluency course experiments that showed changing amounts of UNR also showed changing amounts of the control relative to Ponceau S staining (e.g. Figure 3.4). Whilst a direct link between UNR and γ-tubulin had not been expected at the time, it was nevertheless considered helpful to change the loading control to Ponceau S staining. TCAF1 is involved in promoting the delivery of the TRPM8 channel to the membrane (Gkika et al. 2015). Trpm8 is a cold-activated ion channel, essential for cold sensation in mice (Dhaka et al. 2007) and TRPV2 is an extreme heat-activated ion channel (Caterina et al. 1999). That these proteins are up-regulated following si*UNR* treatment is interesting as it implies a role for UNR in responding to extremes of temperature. Whilst the observation is likely to be coincidental, it is interesting to consider this finding in conjunction with the historic link between cold shock domains and the bacterial cold shock response.

2) t-test p-value less than 0.05, higher in control siRNA samples

In total, there were 104 proteins that had p-values under 0.05 and higher abundances in the control siRNA-treated samples. These may have included multiple versions of a protein with the same gene name, but any such duplicates were later removed prior to carrying out GO-term overrepresentation analysis on the putative hits. The top ten hits that were higher in the control siRNA-treated samples, by t-test p-value, are presented in Table 6.2.

Table 6.2: Top ten proteins with expression levels higher in control siRNA-treated HeLa cells than in si*UNR*-treated HeLa cells, by t-test p-value

Protein	p-value
Prohibitin GN=PHB	5.8E-06
SRA stem-loop-interacting RNA-binding protein, mitochondrial (Fragment) GN=SLIRP	0.00035
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial GN=NDUFA9	0.00074
Actin-like protein 6A GN=ACTL6A	0.00094
Calnexin GN=CANX	0.00103
Voltage-dependent anion-selective channel protein 3 GN=VDAC3	0.00285
Nucleolar complex protein 2 homolog GN=NOC2L	0.00304
DNA-(apurinic or apyrimidinic site) lyase GN=APEX1	0.00346
NADH dehydrogenase (Ubiquinone) flavoprotein 1, 51kDa, isoform CRA_c GN=NDUFV1	0.00368
Dedicator of cytokinesis protein 7 GN=DOCK7	0.00397

The protein potentially down-regulated following si*UNR* treatment with the strongest t-test p-value was prohibitin (Table 6.2). Indeed, this protein had the strongest p-value of any protein differentially regulated between si*UNR* and control siRNA samples recorded in this chapter. Prohibitin is up-regulated in a many cancers as well as being involved in the regulation of mitochondrial membranes, mitochondrial stress and replicative potential (Coates et al. 2001; Coates et al. 1997; Osman et al. 2009). With two of the top hits coming from the ubiquinone complex (NDUFA9 and NDUFV1) and SLIRP also encoding a mitochondrial protein, this implies that UNR may play a role in regulating mitochondrial proteins (Table 6.2). Specifically, it may play a role in the regulation of oxidative phosphorylation. This link to cellular metabolism is interesting when taken in conjunction with the link between UNR and diabetes (Xavier et al. 2014). Such links suggest that UNR may affect cellular respiration at multiple levels from the intake of foodstuffs to the production of ATP.

3) si*UNR*/control siRNA ratio greater than 10

Only five proteins were considered significant on the grounds of having si*UNR*/control siRNA abundance ratios above 10:1 (Table 6.3). None of these were considered significant by t-test p-value (i.e. all p>0.05).

Table 6.3: Proteins with a si*UNR*/control siRNA ratio greater than 10:1

Protein	p-value	si*UNR*/control siRNA
Nucleus accumbens-associated protein 1 GN=NACC1	0.160	∞
Isoform 2 of PITH domain-containing protein 1 GN=PITHD1	0.259	∞
Isovaleryl-CoA dehydrogenase, mitochondrial GN=IVD	0.424	13928.28
Isoform 2 of Apoptosis-inducing factor 2 GN=AIFM2	0.205	41.09
Isoform A of Arfaptin-1 GN=ARFIP1	0.063	10.59

NACC1 was only present in the si*UNR* samples (Table 6.3). This suggests that UNR may act to prevent its expression *in vivo*. NACC1 is an oncogene that represses the expression of GAD45GIP1 (Nakayama et al. 2006; Nakayama et al. 2007). Interestingly, GAD45GIP1 was shown to be somewhat down-regulated on si*UNR* knockdown but the p-value was not considered significant (p=0.268, data not shown).

4) si*UNR*/control siRNA ratio less than 0.1

No proteins had a si*UNR*/control siRNA ratio less than 0.1; the lowest ratio was 0.2152 for Voltage-dependent anion-selective channel protein 3 (Table 6.4).

Table 6.4: Protein with lowest si*UNR*/control siRNA ratio

Protein	p-value	si*UNR*/control siRNA
Voltage-dependent anion-selective channel protein 3 GN=VDAC3	0.0648	0.2152

There was a lack of evidence for the removal of UNR causing any proteins to be down-regulated in terms of having a fold-change greater than ten times (Table 6.4). This suggests that, whilst UNR may modulate the expression of many proteins, it is probably neither essential for their expression nor capable of completely ablating their expression (with the possible exception of the two proteins with infinite ratios in Table 6.3).

6.4 GO-term overrepresentation analyses on *UNR* knockdown-mediated differentially expressed proteins

In order to carry out GO term analyses on the HeLa knockdown data, the proteins that were statistically up-regulated on si*UNR* treatment were added to the 5 proteins that had si*UNR*/control siRNA ratios over 10. Those represented the total number of proteins considered to be up-regulated by si*UNR* treatment. As there were no proteins with si*UNR*/control siRNA ratios under 0.1, those that were statistically down-regulated on si*UNR* treatment were used alone.

6.4.1 GO-term overrepresentation analysis on proteins up-regulated by si*UNR* treatment

The merged protein list was listed by gene name and the single duplicate value was removed, leaving 127 of 128 gene names to be fed into the GO tool at http://amigo.geneontology.org/rte. That returned an unmapped gene and six genes with multiple entries. Acceptable Uniprot identifiers were obtained for these genes and were used to replace them. The gene/protein list was re-entered into the GO tool. The list, as used, is presented in Table 6.5.

Table 6.5: Gene/protein* list corresponding to proteins that are higher in si*UNR*-treated HeLa cells over control siRNA HeLa cells

AADAT	DDX3X	HSP90AB4P	NELFA	PRPS1	SELENBP1	TYMS
ABCF2	DDX5	HSPA14	NFKBIE	PSMC3	SERPINB6	UBE4B
ABCF3	DDX60	HSPA1A	NUBP1	PSMD9	SH3PXD2B	UCHL5
AHCY	DKC1	HSPA1L	P12081	PUS7	SMAD2	UCK2
AHNAK2	DLGAP4	HSPA2	P27361	Q9BY44	SNCG	UGP2
AIFM2	DNAJC7	HSPA8	P40121	Q9H2G2	SSSCA1	UROD
ALDH9A1	DR1	HSPB1	PABPC1	Q9NR30	SUPT6H	YKT6
AP3M1	DYNC1I2	IVD	PARP4	Q9Y3F4	TCAF1	
AP3S1	EIF3I	KIAA1524	PCYT2	RHOA	TIPIN	
ARFIP1	ERVK-5	KIF2A	PEPD	RHOC	TMA7	
C11orf54	FADD	KPNA2	PFN1	RNPEP	TPT1	
C12orf10	FKBP1A	LIMS2	PGD	RPL11	TRAPPC3	
C12orf57	GAPVD1	MAGED2	PGLS	RPL12	TRIM25	
C19orf53	GATAD2A	MOGS	PI4KB	RPL13A	TRPV2	
CBR1	GCA	MRI1	PIR	RPL21	TTC9C	
CDC123	GLO1	MTHFD1	PITHD1	RRM1	TUBB	
CHMP4B	GNE	MYH9	POLR2H	RUFY1	TUBB4A	
CLIC4	HGS	NACC1	PPIA_HUMAN	RXRB	TUBB4B	
CUL1	HMGCS1	NAP1L1	PPM1B	S100A4	TUBG2	
DDX17	HSD17B4	NAP1L4	PRMT5	S100P	TXLNA	

* if gene names were ambiguous or not recognised by the tool, Uniprot identifiers were used.

Brief points concerning two of the genes in Table 6.5:

1) PABPC1 – as stated previously, this protein has its translation repressed by UNR. It is totally expected, therefore, to become up-regulated on *UNR* knockdown (Patel et al. 2005).

2) MYH9 – as was stated in section 4.6.2, MYH9 is associated with invasion and metastasis in osteosarcoma cells (Zhou et al. 2016) and was the highest or second highest protein by total ion intensity observed in the immunoprecipitation data for HeLa and U2OS, respectively (Table 4.6). If si*UNR* treatment in HeLa cells increases the expression of the Myh9 protein, that implies that UNR may repress the expression of the protein, as well as physically binding to it.

The 127 proteins were fed into the GO tool with the following parameters (Table 6.6):

Table 6.6: AmiGO 2 search parameters for all knockdown GO-term analyses

Analysis Type:	PANTHER Overrepresentation Test (release 20160715)
Annotation Version and Release Date:	GO Ontology database Released 2016-09-24
Analyzed List:	upload_1 (Homo sapiens)
Reference List:	Homo sapiens (all genes in database)
Bonferroni correction:	TRUE

The results to this section are to be presented in three sections, by GO term category (biological process, molecular function and cellular component). It should be noted that, whereas the significance-detecting t-tests generated non-multiple testing corrected p-values, the GO tool generated multiple testing corrected p-values using the Bonferroni correction.

1) Overrepresented biological process GO terms using proteins that were more abundant in si*UNR*-treated HeLa cells than in control siRNA-treated cells

There were 21 overrepresented biological process GO terms among those annotated to the proteins in Table 6.5. It was decided to present the top ten over-represented GO terms by p-value and by fold enrichment. The other overrepresented GO terms can be reproduced by entering the gene/Uniprot IDs in Table 6.5 with the parameters in Table 6.6 into the GO tool at http://amigo.geneontology.org/rte.

The top ten over-represented GO terms by p-value are presented in Table 6.7 and the top ten by fold enrichment in Table 6.8.

Table 6.7: Top ten overrepresented biological process GO terms associated with proteins that were higher in abundance in si*UNR* treated HeLa cells, by p-value

GO biological process complete	Ref	Obs	Exp	FE	p-value
viral process (GO:0016032)	831	25	5.03	4.97	2.87E-07
multi-organism cellular process (GO:0044764)	836	25	5.06	4.94	3.26E-07
interspecies interaction between organisms (GO:0044419)	872	25	5.28	4.73	7.86E-07
symbiosis, encompassing mutualism through parasitism (GO:0044403)	872	25	5.28	4.73	7.86E-07
protein refolding (GO:0042026)	20	6	0.12	49.54	2.92E-05
cellular process (GO:0009987)	14523	115	87.95	1.31	7.33E-05
biological_process (GO:0008150)	17008	123	103	1.19	8.13E-04
protein folding (GO:0006457)	227	11	1.37	8	1.39E-03
posttranscriptional regulation of gene expression (GO:0010608)	473	15	2.86	5.24	1.78E-03
positive regulation of viral process (GO:0048524)	105	8	0.64	12.58	2.59E-03

N.B. As in chapter 4, Ref = number of genes with given annotation in the database, Obs = number of genes/proteins observed in experiment, Exp = expected number of genes based on sample size assuming random selection, FE = fold enrichm3ent, p-value as calculated by the GO tool.

Table 6.8: Top ten overrepresented biological process GO terms associated with proteins that were higher in abundance in si*UNR* treated HeLa cells, by fold enrichment

GO biological process complete	Ref	Obs	Exp	FE	p-value
protein refolding (GO:0042026)	20	6	0.12	49.54	2.92E-05
positive regulation of viral life cycle (GO:1903902)	91	7	0.55	12.7	1.36E-02
positive regulation of viral process (GO:0048524)	105	8	0.64	12.58	2.59E-03
positive regulation of multi-organism process (GO:0043902)	161	9	0.97	9.23	6.03E-03
translational initiation (GO:0006413)	154	8	0.93	8.58	4.33E-02
regulation of viral life cycle (GO:1903900)	174	9	1.05	8.54	1.14E-02
protein folding (GO:0006457)	227	11	1.37	8	1.39E-03
posttranscriptional regulation of gene expression (GO:0010608)	473	15	2.86	5.24	1.78E-03
viral process (GO:0016032)	831	25	5.03	4.97	2.87E-07
multi-organism cellular process (GO:0044764)	836	25	5.06	4.94	3.26E-07

Viral process-related GO terms were among the most significant, both by p-value (Table 6.7) and fold enrichment (Table 6.8). Protein folding was the top hit by fold enrichment (Table 6.8).

2) Overrepresented molecular function GO terms using proteins that were more abundant in si*UNR*-treated HeLa cells than in control siRNA-treated cells

There were 34 overrepresented molecular function GO terms among those annotated to the proteins in Table 6.5. The top ten over-represented GO terms by p-value are presented in Table 6.9 and the top ten by fold enrichment in Table 6.10.

Table 6.9: Top ten overrepresented molecular function GO terms associated with proteins that were higher in abundance in si*UNR* treated HeLa cells, by p-value

GO molecular function complete	Ref	Obs	Exp	FE	p-value
protein binding (GO:0005515)	10901	106	66.01	1.61	2.67E-10
cadherin binding involved in cell-cell adhesion (GO:0098641)	277	16	1.68	9.54	4.67E-08
heterocyclic compound binding (GO:1901363)	5953	72	36.05	2	6.50E-08
protein binding involved in cell-cell adhesion (GO:0098632)	287	16	1.74	9.21	7.83E-08
protein binding involved in cell adhesion (GO:0098631)	292	16	1.77	9.05	1.01E-07
cadherin binding (GO:0045296)	295	16	1.79	8.96	1.17E-07
organic cyclic compound binding (GO:0097159)	6037	72	36.56	1.97	1.32E-07
small molecule binding (GO:0036094)	2597	44	15.73	2.8	2.15E-07
nucleotide binding (GO:0000166)	2411	42	14.6	2.88	2.94E-07
nucleoside phosphate binding (GO:1901265)	2412	42	14.61	2.88	2.98E-07

Table 6.10: Top ten overrepresented molecular function GO terms associated with proteins that were higher in abundance in si*UNR* treated HeLa cells, by fold enrichment

GO molecular function complete	Ref	Obs	Exp	FE	p-value
unfolded protein binding (GO:0051082)	106	8	0.64	12.46	8.91E-04
cadherin binding involved in cell-cell adhesion (GO:0098641)	277	16	1.68	9.54	4.67E-08
protein binding involved in cell-cell adhesion (GO:0098632)	287	16	1.74	9.21	7.83E-08
protein binding involved in cell adhesion (GO:0098631)	292	16	1.77	9.05	1.01E-07
cadherin binding (GO:0045296)	295	16	1.79	8.96	1.17E-07
isomerase activity (GO:0016853)	171	9	1.04	8.69	3.16E-03
cell adhesion molecule binding (GO:0050839)	456	16	2.76	5.79	5.52E-05
nucleoside-triphosphatase activity (GO:0017111)	787	18	4.77	3.78	3.75E-03
poly(A) RNA binding (GO:0044822)	1168	26	7.07	3.68	2.03E-05
pyrophosphatase activity (GO:0016462)	832	18	5.04	3.57	8.13E-03

The most significant molecular function GO terms were shown to have stronger Bonferroni-corrected p-values than the most significant biological process GO terms, with the top hit having a p-value 1000 times lower. The tenth most significant molecular function GO term had a similar p-value to that of the top biological process GO term (Table 6.7, Table 6.9). It is possible that UNR regulates the expression of groups of proteins by their molecular function more than by their biological processes. Cadherin binding in cell-cell adhesion was the second highest hit by both p-value and fold enrichment (Tables 6.9 and 6.10). This is interesting in view of the IF work that showed UNR being concentrated at cell-cell interfaces (see section 4.12.1). Poly(A) RNA binding had a fold enrichment of 3.68 and an adjusted p-value of $2x10^{-5}$. This is interesting as UNR is known to interact with a number of mRNAs and to modulate their stability and/or translation. This finding implies that UNR modulates the expression of other proteins that bind to poly(A) regions and are expected, therefore, also to exhibit an effect on translation. Results like that imply that UNR may have a crucial role in the spatiotemporal control of translation.

3) Overrepresented cellular component GO terms using proteins that were more abundant in si*UNR*-treated HeLa cells than in control siRNA-treated cells

There were 36 overrepresented and 2 underrepresented cellular component GO terms among those annotated to the proteins in Table 6.5. The underrepresented GO terms were the closely related ‘intrinsic component of membrane’ (GO:0031224) and ‘integral component of membrane’ (GO:0016021). The top ten over-represented GO terms by p-value are presented in Table 6.11 and the top ten by fold enrichment in Table 6.12.

Table 6.11: Top ten overrepresented cellular component GO terms associated with proteins that were higher in abundance in si*UNR* treated HeLa cells, by p-value

GO cellular component complete	Ref	Obs	Exp	FE	p-value
cytosol (GO:0005829)	3487	69	21.12	3.27	5.23E-19
cytoplasm (GO:0005737)	10781	109	65.29	1.67	3.67E-13
cytoplasmic part (GO:0044444)	8144	93	49.32	1.89	5.07E-12
intracellular part (GO:0044424)	13819	119	83.68	1.42	1.00E-10
extracellular exosome (GO:0070062)	2735	49	16.56	2.96	6.07E-10
extracellular vesicle (GO:1903561)	2749	49	16.65	2.94	7.36E-10
extracellular organelle (GO:0043230)	2750	49	16.65	2.94	7.46E-10
intracellular (GO:0005622)	14192	119	85.94	1.38	1.57E-09
adherens junction (GO:0005912)	684	24	4.14	5.79	5.06E-09
anchoring junction (GO:0070161)	702	24	4.25	5.65	8.66E-09

Table 6.12: Top ten overrepresented cellular component GO terms associated with proteins that were higher in abundance in si*UNR* treated HeLa cells, by fold enrichment

GO cellular component complete	Ref	Obs	Exp	FE	p-value
cell-cell adherens junction (GO:0005913)	320	16	1.94	8.26	1.84E-07
cytosolic part (GO:0044445)	243	9	1.47	6.12	2.52E-02
adherens junction (GO:0005912)	684	24	4.14	5.79	5.06E-09
anchoring junction (GO:0070161)	702	24	4.25	5.65	8.66E-09
focal adhesion (GO:0005925)	393	12	2.38	5.04	7.09E-03
cell-substrate adherens junction (GO:0005924)	396	12	2.4	5	7.65E-03
cell-substrate junction (GO:0030055)	400	12	2.42	4.95	8.46E-03
cell-cell junction (GO:0005911)	641	18	3.88	4.64	9.42E-05
cytosol (GO:0005829)	3487	69	21.12	3.27	5.23E-19
cell junction (GO:0030054)	1371	26	8.3	3.13	2.37E-04

It is clear that the cellular component GO terms have the strongest p-values of the three groups (Table 6.11 versus Tables 6.7 and 6.9). That said, the top four hits are extremely general and more important are possibly the extracellular exosome (as it has been seen previously [e.g. Table

4.12C]) and adherens junction (as it had also been seen previously and it has a much smaller reference list [e.g. Table 4.19D, Table 6.12]). The cell-cell adherens junction had the greatest fold enrichment (8.26) and also had a strong adjusted p-value ($p=1.84x10^{-7}$).

6.4.2 Consideration of GO-term overrepresentation analysis on proteins present at higher levels in si*UNR*-treated HeLa cells than in control siRNA-treated HeLa cells

It appears that UNR could possibly be regulating the expression of groups of proteins based on their cellular location more than the biological process in which they are involved. It is not immediately obvious why that should be the case. It is possible that UNR is involved in the sequestration of mRNA. The removal of UNR could then either release the mRNA to be translated and the resultant protein to be transported to its site of action, or the transcript itself could be localised to a specific region to be translated locally. Another possibility is that UNR is itself localised within the cell and either increases or decreases the translation of transcripts concentrated in those areas. However, a localised concentration of UNR in specific regions of the cell under all conditions is not in keeping with IF work carried out in the Anderson lab over the past few years (data not shown). A third possibility is that UNR aids somehow in the degradation of mRNA or protein. It is also possible that UNR exerts an effect on translation that exists outside a three dimensional snapshot of the cell. By controlling the translation of some transcription factors, UNR could affect the transcription of groups of transcripts that are involved in specific functions at specific regions of the cell. Also, by controlling the translation of a large number of poly(A) binding proteins (at least 26 – Table 6.10), UNR could control translation generally across almost all transcripts whilst also binding to and affecting the translation of a smaller subset of transcripts directly. The top molecular function GO terms are intermediate in p-value between the biological process GO terms and the cellular component GO terms.

6.4.3 GO-term overrepresentation analysis on proteins down-regulated by si*UNR* treatment

The merged protein lists were listed by gene name and any duplicate values were removed. These were then initially fed into the GO tool at http://amigo.geneontology.org/rte. That

returned two genes with multiple entries. Acceptable Uniprot identifiers were obtained for those genes and were used to replace them. The gene/protein list was re-entered into the GO tool. The 104 entry list, as used, is presented in Table 6.13.

Table 6.13: Gene/protein* list corresponding to proteins that are higher in control siRNA-treated HeLa cells over si*UNR*-treated HeLa cells

ABCE1	CDH10	EXOSC5	LAMB3	PTPN1	TOMM70A
ACOT13	CELF1	FAM192A	MBOAT7	PTRH2	TOR1AIP1
ACTL6A	CISD1	FTH1	MPDU1	Q9H910	TST
ADD3	COA7	FUNDC2	MTCH2	RAB35	TTC7B
AGO2	COX4I1	FUS	MYO6	RANBP9	UBR4
AGTPBP1	COX5A	GALNT2	NDRG1	RBM14	UGGT1
AK2	COX6B1	GALNT7	NDUFA9	RCC2	UQCRC2
ALG11	CPOX	GLOD4	NDUFS5	SDHA	VAPA
APEX1	CSDE1	GSTK1	NDUFV1	SEC62	VDAC1
ATG5	CTNNA2	HDAC3	NOC2L	SLC25A3	VDAC3
ATP2A2	CYC_HUMAN	HDLBP	NOTCH2	SLC2A14	WDR33
ATP2A3	DDX1	HIBADH	NR3C1	SLIRP	WRNIP1
ATP5L	DOCK7	HSPE1	P31947	SNRPB2	YWHAG
BNIP3	DSG2	IGF2R	PDCD4	SQRDL	YWHAH
C8orf82	ECH1	ITGB1	PDIA6	STAT6	
CANX	EHD4	JUP	PHB	STT3A	
CCDC51	EPHX1	KHSRP	PPP4R2	TMEM194A	
CCDC58	ERGIC3	KIAA0319L	PRRC2A	TOMM40	

* if gene names were ambiguous, appropriate Uniprot identifiers were entered into the tool.

Fortunately, UNR was shown to be down-regulated following *UNR* knockdown (Table 6.13). Another gene that seemed important on the author's uneducated cursory scan of Table 6.13 was AGO2 (i.e. there may be much more important genes in the list with functions with which the author is unfamiliar). Argonaut-2 is responsible for the RISC-mediated cleavage of mRNA following siRNA treatment (Matranga et al. 2005). This was considered interesting because it was down-regulated on si*UNR* treatment over control siRNA treatment, implying that UNR may be involved in the expression of AGO2.

As with the previous GO term analyses, these results are presented in three sections by GO term category.

1) Overrepresented biological process GO terms using proteins that were more abundant in control siRNA-treated HeLa cells than in si*UNR*-treated cells

There were 24 overrepresented biological process GO terms among those annotated to the proteins in Table 6.13. The top ten over-represented GO terms by p-value are presented in Table 6.14 and the top ten by fold enrichment in Table 6.15.

Table 6.14: Top ten overrepresented biological process GO terms associated with proteins that were higher in abundance in control siRNA treated HeLa cells, by p-value

GO biological process complete	Ref	Obs	Exp	FE	p-value
oxidative phosphorylation (GO:0006119)	98	9	0.49	18.52	1.62E-05
respiratory electron transport chain (GO:0022904)	108	9	0.54	16.8	3.74E-05
electron transport chain (GO:0022900)	111	9	0.55	16.35	4.73E-05
mitochondrial ATP synthesis coupled electron transport (GO:0042775)	88	8	0.44	18.33	1.44E-04
ATP synthesis coupled electron transport (GO:0042773)	89	8	0.44	18.13	1.57E-04
mitochondrial transport (GO:0006839)	186	10	0.92	10.84	3.05E-04
ATP metabolic process (GO:0046034)	204	10	1.01	9.88	7.15E-04
mitochondrion organization (GO:0007005)	632	16	3.13	5.11	8.20E-04
cellular respiration (GO:0045333)	164	9	0.81	11.07	1.28E-03
purine ribonucleoside triphosphate metabolic process (GO:0009205)	227	10	1.13	8.88	1.89E-03

Table 6.15: Top ten overrepresented biological process GO terms associated with proteins that were higher in abundance in control siRNA treated HeLa cells, by fold enrichment

GO biological process complete	Ref	Obs	Exp	FE	p-value
mitochondrial electron transport, cytochrome c to oxygen (GO:0006123)	20	4	0.1	40.33	2.92E-02
oxidative phosphorylation (GO:0006119)	98	9	0.49	18.52	1.62E-05
mitochondrial ATP synthesis coupled electron transport (GO:0042775)	88	8	0.44	18.33	1.44E-04
ATP synthesis coupled electron transport (GO:0042773)	89	8	0.44	18.13	1.57E-04
respiratory electron transport chain (GO:0022904)	108	9	0.54	16.8	3.74E-05
electron transport chain (GO:0022900)	111	9	0.55	16.35	4.73E-05
cellular respiration (GO:0045333)	164	9	0.81	11.07	1.28E-03
mitochondrial transport (GO:0006839)	186	10	0.92	10.84	3.05E-04
ATP metabolic process (GO:0046034)	204	10	1.01	9.88	7.15E-04
purine ribonucleoside triphosphate metabolic process (GO:0009205)	227	10	1.13	8.88	1.89E-03

Whilst many of the GO terms shared the same proteins, it appeared that the top biological process GO terms overrepresented on UNR downregulation with si*UNR* were related to oxidative phosphorylation and closely related mitochondrial GO terms. This was both in terms of p-value (Table 6.14) and fold enrichment (Table 6.15).

2) Overrepresented molecular function GO terms using proteins that were more abundant in control siRNA-treated HeLa cells than in si*UNR*-treated cells

There were 8 overrepresented molecular function GO terms among those annotated to the proteins in Table 6.13. All 8 over-represented GO terms by p-value are presented in Table 6.16. As all overrepresented GO terms are presented in Table 6.16, there was no reason to include a separate fold enrichment table.

Table 6.16: Overrepresented molecular function GO terms associated with proteins that were higher in abundance in control siRNA treated HeLa cells, by p-value

GO molecular function complete	Ref	Obs	Exp	FE	p-value
protein binding involved in cell-cell adhesion (GO:0098632)	287	12	1.42	8.43	6.19E-05
protein binding involved in cell adhesion (GO:0098631)	292	12	1.45	8.29	7.46E-05
cadherin binding involved in cell-cell adhesion (GO:0098641)	277	11	1.37	8.01	4.10E-04
cadherin binding (GO:0045296)	295	11	1.46	7.52	7.62E-04
cell adhesion molecule binding (GO:0050839)	456	12	2.26	5.31	8.05E-03
RNA binding (GO:0003723)	1631	23	8.09	2.84	1.11E-02
porin activity (GO:0015288)	6	3	0.03	> 100	1.11E-02
poly(A) RNA binding (GO:0044822)	1168	18	5.79	3.11	4.60E-02

It was noted that there were fewer significant molecular function GO terms using the proteins down-regulated following si*UNR* treatment compared to those found using up-regulated proteins. It was noted that the same general themes were present with both the up-regulated and down-regulated proteins, however. For example, cadherin binding, cell-cell adhesion and poly(A) RNA binding GO terms are present in both Table 6.10 and Table 6.16. It was noted that the p-values for these were much stronger with the up-regulated proteins than the down-regulated ones. This could in part be due to there being more proteins in the up-regulated list (Table 6.5 versus Table 6.13). Whatever the reason, however, it appears that *UNR* knockdown both increases and decreases the level of proteins with given molecular function GO term annotations. This is to be expected if UNR exerts an effect on whole molecular function GO terms and, as discussed at length in the introduction, UNR is already known to promote the translation of certain mRNAs and repress the translation of others.

3) Overrepresented cellular component GO terms using proteins that were more abundant in control siRNA-treated HeLa cells than in si*UNR*-treated cells

There were 58 overrepresented cellular component GO terms among those annotated to the proteins in Table 6.13. The top ten over-represented GO terms by p-value are presented in Table 6.17 and the top ten by fold enrichment in Table 6.18.

Table 6.17: Top ten overrepresented cellular component GO terms associated with proteins that were higher in abundance in control siRNA treated HeLa cells, by p-value

GO cellular component complete	Ref	Obs	Exp	FE	p-value
intracellular membrane-bounded organelle (GO:0043231)	10985	99	54.47	1.82	4.84E-19
mitochondrion (GO:0005739)	1717	45	8.51	5.29	7.22E-19
membrane-bounded organelle (GO:0043227)	12122	100	60.11	1.66	3.10E-16
intracellular organelle part (GO:0044446)	8167	85	40.5	2.1	5.88E-16
intracellular organelle (GO:0043229)	12049	99	59.75	1.66	2.63E-15
organelle part (GO:0044422)	8351	85	41.41	2.05	2.98E-15
cytoplasmic part (GO:0044444)	8144	84	40.39	2.08	3.23E-15
organelle envelope (GO:0031967)	1145	33	5.68	5.81	8.06E-14
envelope (GO:0031975)	1151	33	5.71	5.78	9.38E-14
mitochondrial part (GO:0044429)	998	31	4.95	6.26	1.17E-13

Table 6.18: Top ten overrepresented cellular component GO terms associated with proteins that were higher in abundance in control siRNA treated HeLa cells, by fold enrichment

GO cellular component complete	Ref	Obs	Exp	FE	p-value
intercalated disc (GO:0014704)	53	5	0.26	19.02	9.94E-03
respiratory chain complex (GO:0098803)	81	7	0.4	17.43	2.53E-04
organelle envelope lumen (GO:0031970)	85	7	0.42	16.61	3.49E-04
mitochondrial respiratory chain (GO:0005746)	87	7	0.43	16.23	4.08E-04
respiratory chain (GO:0070469)	100	8	0.5	16.13	5.92E-05
mitochondrial intermembrane space (GO:0005758)	77	6	0.38	15.71	3.53E-03
cell-cell contact zone (GO:0044291)	66	5	0.33	15.28	2.83E-02
inner mitochondrial membrane protein complex (GO:0098800)	130	8	0.64	12.41	4.27E-04
mitochondrial protein complex (GO:0098798)	163	10	0.81	12.37	1.39E-05
mitochondrial membrane part (GO:0044455)	204	12	1.01	11.86	7.00E-07

The top ten overrepresented cellular component GO terms for the proteins down-regulated by si*UNR* treatment were quite different to those using up-regulated proteins (Tables 6.17 and 6.11). Nevertheless, there were 58 significant cellular component GO terms found using the down-regulated proteins. Those 58 GO terms were listed with the top ten cellular component GO terms by p-value for the up-regulated proteins and Microsoft Excel was used to find

duplicates. This showed that nine out of the ten GO terms in Table 6.11 (all except cytosol) were also considered significantly overrepresented among si*UNR*-mediated down-regulated proteins (Bonferroni corrected p-value<0.05 - the actual p-value range was 0.000816 to 0.00999). The data in Tables 6.17 and 6.18 do seem reliable as the p-values in them are strong. It is possible that UNR controls the expression of different functional groups of proteins by generally promoting their expression or generally inhibiting their expression.

6.4.4 Consideration of GO-term overrepresentation analysis on proteins down-regulated by si*UNR* treatment

The si*UNR*-mediated knockdown of *UNR* in HeLa cells resulted in approximately the same number of significantly up-regulated (127, Table 6.5) and down-regulated (104, Table 6.13) proteins by the criteria set out in the text. As with the up-regulated protein data, the most significantly overrepresented biological process and molecular function GO terms had weaker p-values than the most significant cellular component GO terms. The most significant GO term p-values were 1.62×10^{-5}, 6.19×10^{-5} and 4.84×10^{-19}, respectively. There were more significant mitochondrion-related biological process and cellular component GO terms with the down-regulated proteins compared to the up-regulated proteins, for example the 'mitochondrion' GO term had an adjusted p-value of 7.22×10^{-19} (Table 6.17). It is possible therefore that UNR has a role in the promotion of mitochondrial protein expression as knocking down UNR downregulates the proteins. Further work would be needed to see if UNR acts on translation of mitochondrial genes directly, either in the mitochondrion or in the cytoplasm for proteins targeted to the mitochondrion. It is also possible that UNR exerts its control on mitochondrial protein levels indirectly, for example by modulating the expression of transcription factors or by physically binding to other proteins (see section 6.6.3).

6.5 Knockdown of *UNR* in U2OS cells

Two sets of three repeats were carried out for the *UNR* knockdown experiments in U2OS cells. One set was stressed with 1 mM sodium arsenite for 1 hour prior to harvesting whereas the other set were mock treated with the same volume of sterile PBS.

6.5.1 Western blot confirmation of successful knockdown

Approximately one twentieth of the cell lysates was retained for analysis by Western blot in order to confirm that the knockdown had been successful. In the case of the U2OS samples, all the unstressed U2OS knockdowns clearly worked as expected although most of the tubulin blots were a bit out of focus (Figure 6.4). The second and third stressed U2OS samples also worked as expected but there was very little staining for UNR in the first control sample (Figure 6.4). It was decided to proceed to process all of the plus arsenite samples, however, as there did appear to be more UNR in lane stressed C1 than in lane stressed U1 even though there was less tubulin in in lane stressed C1 than in lane stressed U1 (Figure 6.4). On top of this, it was noted that there were around 30% fewer overall proteins detected in the unstressed C1 samples compared to the other 5 unstressed samples (Figure 6.5). Whilst there could be differences between the stressed and unstressed samples, there would not be expected to be extreme differences in the total number of proteins observed and any differences would be expected to occur over all samples. As it was, there were 31.4% fewer proteins detected by Scaffold in the first unstressed control U2OS sample compared to the first unstressed si*UNR* sample. Compared to this, the equivalent differences for the other two repeats were 5.1% more proteins (repeat 2) and 3.7% more proteins (repeat 3) (Figure 6.5). This led to the assumption that there were fewer overall proteins in the first repeat arsenite-stressed control siRNA-treated sample. Whilst some proteins would not be present in that control sample that were present in the other repeats, it was hoped that Progenesis' normalisation function would make those that were present adequately comparable with the abundance present in the equivalent si*UNR* sample.

Due to the unconventional presentation of Figure 6.4, it is worth pointing out exactly what it shows. Additional knockdowns of *TP53* and a combination of *UNR* and *TP53* were between each set of si*UNR* and control samples. As there will be no discussion of the additional knockdowns in this thesis, it was considered helpful to block out those lanes with orange rectangles to make the pertinent samples more obvious. At the same time, it was considered useful to show the initial images from which the altered images were made. The fourth well from the left was blank; this can be best seen in the bottommost tubulin panel and the drawn-on red and blue marks to the left of centre refer to bands from a marker than was run in that lane.

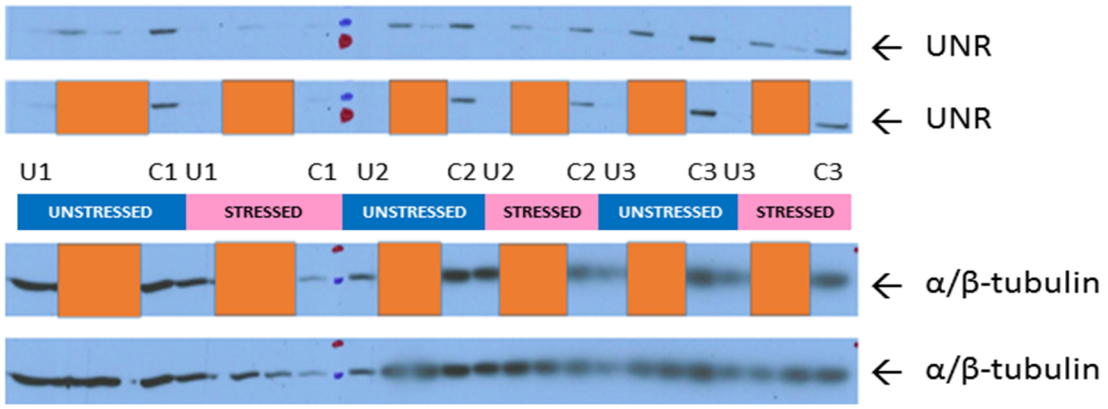


Figure 6.4: Western blot showing *UNR* knockdown. U2OS cells were treated with si*UNR* or a control siRNA (as stated) 48 hours prior to harvesting. One hour prior to harvesting, the cells were fed with fresh growth medium with (stressed, pink) or without (unstressed, blue) 1 mM sodium arsenite. The images show UNR (upper two panels) or α/β-tubulin (lower two panels). U = si*UNR* sample, C = control siRNA sample, numbers = repeat. The top two panels are identical, as are the lower two. The orange boxes in the middle panels hide bands that are visible in the other panels but do not form part of the discussion. The red and blue marks on the panels refer to protein size markers (upper markers: red = 70kDa, blue = 100kDa; lower markers: red = 70kDa, blue = 55kDa).

6.5.2 Error in processing the unstressed U2OS samples in Progenesis

Due to a labelling error in processing the unstressed U2OS samples in Progenesis, compounded by its late detection and loss of access rights to Life Sciences, a decision had to be made between presenting Progenesis analysis of two of the three sets of repeats based upon normalisation calculations that included an incorrect sample, and reanalysing the unstressed samples using Scaffold. It was decided to use Scaffold.

6.5.3 Justification of the use of Scaffold

Looking at the immunoprecipitation data from unstressed HeLa cell lysates, it was clear that the Scaffold hits and the Progenesis hits were similar, although there were more hits with Progenesis. Taking the 13 proteins listed in Table 4.2B (Scaffold data), seven are listed in the equivalent Progenesis analysis (Table 4.7B) (UNRIP, NARR, LDB1, UNR, SQSTM1, LMO4 and TFG). One further protein is listed as having a high UNR/IgG ratio in Table 4.9B (MYCBP2). Two further proteins were listed as having been pulled down with UNR in arsenite-treated HeLa cells

(HUWE1 and RPS5) (Table 4.8B). The SSBP4 protein was not found in the Progenesis data, but both the SSBP2 and SSBP3 proteins were (Table 4.7B). Using the protein alignment tool at www.uniprot.org on 13/10/16, SSBP4 showed 69.5% identity with SSBP2, 71.0% identity with SSBP3 and 62.7% identity was retained across all three. It may be the case that these proteins were confused between the two programs. Of the other two proteins present in Table 4.2B but not in Table 4.7B, HSPD1 showed a distribution that questioned whether or not it was a genuine hit. The largest value in the IgG repeats (130270) was larger than that of two of the UNR values (40965 and 34231). Whilst there was more of the protein present in every UNR sample over the equivalent IgG sample, the large amount that was pulled down by IgG on one occasion suggested that HSPD1 may be false positive.

Both MYCBP2 and HUWE1 had protein distributions that would suggest that they were true positives as all UNR pulldown values were at least an order of magnitude greater than the corresponding IgG pulldowns and, in the case of HeLa5 versus HeLa2 for HUWE1, more than two orders of magnitude (Table 6.19). That, and the fact that HUWE1 and RPS5 had the same p-value even though the IgG pulldown values for RPS5 were much closer (ratios= 1:1.08, 1:1.09, 1:2.10, Table 6.19), supports the idea that t-tests are not an ideal way to decide on significance.

Table 6.19: t-test p-values using Progenesis-calculated protein levels for selected proteins considered significant by Scaffold but not by Progenesis

Protein	HeLa IgG_1	HeLa IgG_2	HeLa IgG_3	HeLa Unr_1	HeLa Unr_2	HeLa Unr_3	p-value
MYCBP2	10,756	14,613	11,058	422,964	192,383	137,694	0.1130
HUWE1	859	514	11,976	75,310	118,533	430,660	0.2005
RPS5	461,654	630,056	189,952	499,451	688,074	398,701	0.2005

Due to the fact that most putative hits using Scaffold were also flagged up by Progenesis and that at least some of those that were not nevertheless appear to be valid, it was considered acceptable to analyse the knockdown data using Scaffold and paired t-tests alone. The advantages of Progenesis were still acknowledged, and the use of Scaffold was only intended to give an initial general insight into the identity of proteins and groups of proteins that may be affected by *UNR* knockdown, purely because circumstances prevented equivalent analysis with Progenesis.

6.6 Analysis of data obtained from U2OS cells following si*UNR* or a control siRNA treatment, without exogenous stress

The raw mass spectrometry data from the lysates made from unstressed U2OS cells that had been treated with si*UNR* or a control siRNA were fed into Scaffold by the proteomics department of Warwick University. The publication report for this was exactly the same as that in section 4.3.1. Scaffold was set to give quantitative output as normalised total ion current. There were 1917 protein clusters identified across the six samples and, with the exception of the first control siRNA repeat, there were approximately the same number of proteins identified and spectra detected for each sample (Figure 6.5).

Category	Bio Sample	MS/MS Sample	#Prot	#IDs	#Spec	%IDs
-siUnr	-siUnr1	OT_160204_5 (F009811)	1398	9093	41134	22%
-siRNA	-siRNA 1	OT_160204_11 (F009814)	959	5614	31876	18%
-siUnr	-siUnr2	OT_160204_13 (F009815)	1309	7056	31175	23%
-siRNA	-siRNA 2	OT_160204_19 (F009818)	1376	10574	52109	20%
-siUnr	-siUnr3	OT_160204_21 (F009819)	1341	7993	40618	20%
-siRNA	-siRNA 3	OT_160204_27 (F009822)	1390	8746	43061	20%

Figure 6.5: Scaffold-derived table showing the number of spectra per sample (#Spec) and number of proteins detected under the parameters laid out in the text (#Prot).

6.6.1 Detection of proteins that are differentially expressed in U2OS cells following treatment with si*UNR* or a control siRNA, without exogenous stress

The Scaffold data were then exported to Microsoft Excel for further analysis. As in chapter 4, a liberal two-tailed paired t-test p-value significance level of $p<0.1$ was used. This was used to increase the number of significant proteins detected by t-test to make it more comparable to the number detected from the HeLa samples. That was, in turn, to make the p values obtained from the GO tool more comparable between the HeLa and U2OS samples. It also made the Scaffold analysis more comparable with the Scaffold analyses carried out in Chapter 4, which were discussed in section 6.5.3 in respect to validating the use of Scaffold in this section). Proteins that were at least ten times more abundant in one of the conditions (si*UNR* or control) were also included. Where the si*UNR*/control siRNA ratio was 0 or infinite, the protein was included only when it was present in all three of the repeats for the condition in which it was detected. Lists were made for those proteins that went up on si*UNR* by t-test p-value or ratio and duplicated proteins were removed after the lists were merged. This was repeated for

proteins that were down-regulated following si*UNR* treatment. Finally, both lists were merged to give a list of proteins that were differentially expressed following treatment with si*UNR*.

6.6.2 Discovery of proteins with significantly different expression levels in U2OS cells following treatment with si*UNR*, without exogenous stress

The changes in protein levels were explored in turn for each of the four groups used for the HeLa samples.

1) t-test p-value less than 0.1, higher in si*UNR* samples

There were 100 proteins with a p-value less than 0.1 that were higher in the si*UNR* samples. The top ten of these are presented in Table 6.20.

Table 6.20: Top ten proteins whose expression is up-regulated following si*UNR* treatment in unstressed U2OS cells, by t-test p-value

Protein	p-value
Nuclease EXOG, mitochondrial GN=EXOG	0.00227
Isoform 3 of Proteasome subunit beta type-5 GN=PSMB5 (P28074-3)	0.00440
Transcription factor A, mitochondrial GN=TFAM	0.00554
E3 SUMO-protein ligase RanBP2 GN=RANBP2	0.00626
Serine/threonine-protein phosphatase 1 regulatory subunit 10 GN=PPP1R10	0.00668
A-kinase anchor protein 1, mitochondrial GN=AKAP1	0.00728
Interferon regulatory factor 2-binding protein 2 GN=IRF2BP2	0.00729
G protein-regulated inducer of neurite outgrowth 1 GN=GPRIN1	0.00975
Cytochrome c oxidase assembly protein COX19 GN=COX19	0.01064
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11, mitochondrial GN=NDUFB11	0.01216

2) t-test p-value less than 0.1, higher in control siRNA samples

There were 53 proteins with a p-value less than 0.1 that were higher in the control siRNA samples. The top ten of these are presented in Table 6.21.

Table 6.21: Top ten proteins whose expression is lower following si*UNR* treatment in unstressed U2OS cells, by t-test p-value

Protein	p-value
Nucleolysin TIA1 isoform p40 GN=TIA1	0.00027
Prosaposin GN=PSAP	0.00522
Putative G antigen family E member 3 GN=PAGE2B	0.00683
Nectin-3 GN=PVRL3	0.00706
AP2-associated protein kinase 1 (Fragment) GN=AAK1	0.00921
Proteasome subunit beta type-3 GN=PSMB3	0.01066
Density-regulated protein GN=DENR	0.01109
Microtubule-associated protein GN=MAP4	0.01129
Poly(rC)-binding protein 1 GN=PCBP1	0.01282
MICOS complex subunit MIC19 GN=CHCHD3	0.01438

3) si*UNR*/control siRNA ratio greater than 10

There were 19 proteins that were at least ten times higher in the si*UNR* samples. Of these, 14 were only observed in the si*UNR* samples (Table 6.22). It should be reiterated here that proteins with zero or infinite ratios were only included when they had zero abundance in all samples from one group and non-zero abundances in all samples from the other group.

Table 6.22: Proteins observed in si*UNR* samples only

Zinc finger protein 106 GN=ZNF106
Isoform 2 of Suprabasin GN=SBSN (Q6UWP8-2)
Protein NipSnap homolog 3A GN=NIPSNAP3A
Ferritin GN=FTH1
Vitamin K epoxide reductase complex subunit 1-like protein 1 GN=VKORC1L1
Nuclease EXOG, mitochondrial GN=EXOG
Transcription factor A, mitochondrial GN=TFAM
RNA binding motif protein 10, isoform CRA_d GN=RBM10
Serine/threonine-protein phosphatase 1 regulatory subunit 10 GN=PPP1R10
Dipeptidyl peptidase 1 GN=CTSC
Fibronectin GN=FN1
Cytochrome c oxidase assembly protein COX19 GN=COX19
Isoform 6 of RNA binding protein fox-1 homolog 2 GN=RBFOX2 (O43251-6)
SAFB-like transcription modulator GN=SLTM

4) control siRNA/si*UNR* ratio greater than 10

There were four proteins that were present in the control siRNA samples more than ten times more than in the si*UNR* samples (Table 6.23).

Table 6.23: Proteins with a control siRNA/si*UNR* ratio greater than 10:1

Protein	control siRNA/siunr ratio
MICOS complex subunit MIC19 GN=CHCHD3	∞
Glutathione S-transferase P GN=GSTP1	15.52
Lactoylglutathione lyase GN=GLO1	13.69
Translin-associated protein X GN=TSNAX	11.88

When the proteins (or protein clusters) with significantly greater abundances in the si*UNR*-treated samples by t-test p-value were merged with those at least ten times more abundant in the si*UNR*-treated samples and duplicate gene names were removed, 101 entries were left.

MIC19 was present in both Table 6.21 and Table 6.23, but the other three proteins that were significantly lower in the si*UNR* samples by ratio were not significant by p-value. 56 proteins (or protein clusters) were therefore accepted as being significantly lower in the si*UNR* samples (i.e. 53 by p value and three by ratio).

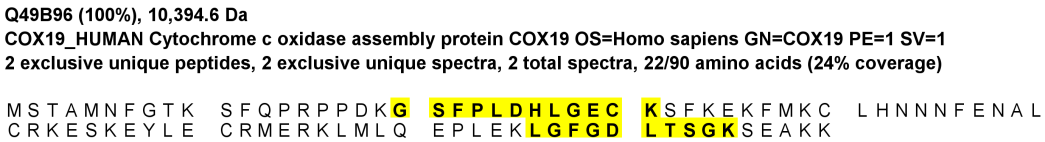
6.6.3 Consideration of token proteins that were differentially expressed between si*UNR*-treated and control siRNA-treated U2OS cells, without exogenous stress

1) Proteins higher in si*UNR* samples

With respect to the musings in section 6.4.4 as to the possibility of UNR controlling the expression of mitochondrial proteins, the third most significantly up-regulated protein following *UNR* knockdown in unstressed U2OS cells, by t-test p-value, was TFAM (Table 6.20). That protein is a mitochondrial transcription factor (Parisi& Clayton 1991). Only seeing it upon *UNR* knockdown would imply that UNR may have a role in supressing its expression. This observation

may merit future investigation as a result of it offering a potential route by which UNR can modulate the expression of mitochondrial proteins. Other proteins in the top ten up-regulated following si*UNR* treatment included a mitochondrial AKAP, COX19, a ubiquinone subunit and the top hit was the mitochondrial nuclease, EXOG (Table 6.20). This implies that the removal of UNR may preferentially upregulate mitochondrial proteins in U2OS cells. Looking at one of these proteins, COX19, showed that two exclusive unique peptides for the protein, equating to 24% coverage, had been observed (Figure 6.6A). The protein was observed in all si*UNR* samples but in none of the control siRNA samples (Figure 6.6B).

A



B

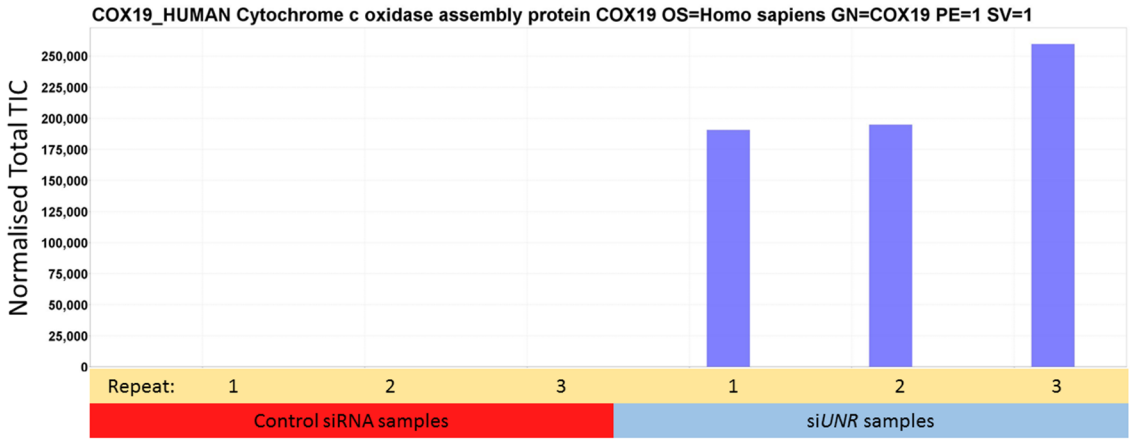


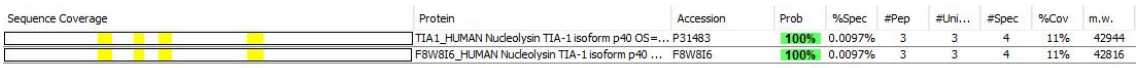
Figure 6.6: Graphical exploration of the validity of COX19 as a protein with higher abundance in unstressed U2OS cells following treatment with si*UNR*. (A) – primary sequence of the protein with yellow shading showing the observed peptides, (B) – bar chart showing the amount of observed COX19 in different samples, as stated. Images were taken from the Scaffold software and (B) was altered to make its text clearer.

2) Proteins higher in control siRNA samples

Among the proteins that were lower in the si*UNR*-treated samples was the top hit, TIA1 (Table 6.21). TIA1 aggregates form stress granules and it has been shown here, in the literature (White

& Lloyd 2011), and by other work in the Anderson lab that UNR is found in stress granules in certain arsenite stressed cells. Stress granules contain non-translating mRNAs and can offer protection to the transcripts whilst changing the translational output through their alternate sequestration and liberation (Buchan & Parker 2009). If TIA1 levels decrease on *UNR* knockdown, this implies that UNR may play a role in promoting the translation of TIA1. The TIA1 protein has two forms that vary by the presence or absence of a glutamine residue (QQ→QQQ) and Scaffold assigned peptides to the shorter (F8W8I6) version to maximise coverage. The peptide coverage is otherwise exactly the same (Figure 6.7A). There were four peptides and three exclusive unique peptides recorded for TIA1, equating to 11% coverage (Figure 6.7B). The control si*UNR* samples all contained approximately twice the level of TIA1 observed in their associated si*UNR* samples (Figure 6.7C).

A



B

F8W8I6 (100%), 42,834.1 Da
F8W8I6_HUMAN Nucleolysin TIA-1 isoform p40 OS=Homo sapiens GN=TIA1 PE=1 SV=1
3 exclusive unique peptides, 3 exclusive unique spectra, 4 total spectra, 43/385 amino acids (11% coverage)

MEDEMPKTLY VGNLSRDVTE ALILQLFSQI GPCKNCKMIM DTAGNDPYCF
VEFHEHRHAA AALAAMNGRK IMGKEVKVNW ATTPSSQKKD TSSSTVVSTQ
RSQDHFHVFV GDLSPEITTE DIKAAFAPFG RISDARVVKD MATGKSKGYG
FVSFFNKWDA ENAIQQMGGQ WLGGRQIRTN WATRKPPAPK STYESNTKQL
SYDEVVNQSS PSNCTVYCGG VTSGLTEQLM RQTFSPFGQI MEIRVFPDKG
YSFVRFNSHE SAAHAIVSVN GTTIEGHVVK CYWGKETLDM INPVQQNQIG
YPQPYGQWGQ WYGNAQQIGQ YMPNGWQVPA YGMYGQAWNQ QGFNQTQSSA
PWMGPNYGVQ PPQGQNGSML PNQPSGYRVA GYETQ

C

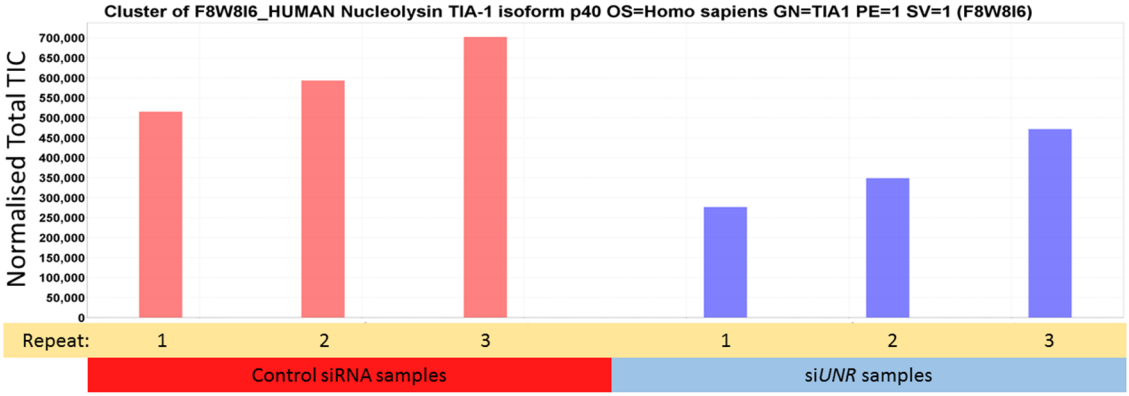


Figure 6.7: Graphical exploration of the validity of TIA1 as a protein with lower abundance in unstressed U2OS cells following treatment with si*UNR*. (A) – schematic diagram showing the coverage of TIA1 and F8W8I6, (B) – the primary sequence of the F8W8I6 version of TIA1 with yellow shading showing the observed peptides, (C) – bar chart showing the amount of observed COX19 in different samples, as stated. Images were taken from the Scaffold software and (C) was altered to make its text clearer.

6.6.4 Putative hits exported to the AMIGO2 GO-term overrepresentation tool

The full list of putative si*UNR*-mediated up- or down-regulated proteins were fed into the GO tool at http://amigo.geneontology.org/rte. The gene names were obtained as before and ambiguous or unrecognised terms were replaced with Uniprot identifiers where possible. The

gene *DKFZp566H192* was found to encode the protein neuroplastin (Q9UFM8) which was up-regulated following si*UNR* treatment. That said, neither the gene name nor the Uniprot ID were recognised by the GO tool, so the gene/protein was removed. Likewise, the down-regulated but Uniprot-uncharacterised protein A0A087WV05 was removed.

These alterations resulted in 100 proteins that were more abundant in unstressed si*UNR*-treated U2OS cells and 55 proteins that were more abundant in unstressed control siRNA-treated U2OS cells being used for GO term overrepresentation analysis. The full lists, as used, are presented in Table 6.24.

Table 6.24: Gene names* of proteins differentially regulated following si*UNR* treatment

A: Proteins higher in si*UNR*-treated samples

AHNAK2	CLCC1	FN1	ITGB5	PALM	SCRIB	TP53BP1
AKAP1	CNPY3	FTH1	KAT7	PCM1	SERBP1	TPD52L2
AKAP12	COX19	GATAD2B	KTN1	POLR2G	SGTA	VDAC3
ALDH2	CPD	GGCT	LBR	PPP1R10	SLAIN2	VKORC1L1
ARPC2	CTSC	GIGYF2	LEMD3	PRPF3	SLC30A1	WBP11
BAD	DAB2	GLRX5	LETM1	PSMB5	SLC44A1	YLPM1
BCAM	DCBLD2	GPRIN1	LGALS3	Q8IWI9	SLC9A3R2	ZC3H11A
BPTF	DECR1	HIST1H1T	MRPS36	RANBP2	SLTM	ZNF106
BSG	EEA1	HMGN5	MT-ATP8	RBFOX2	SRRM2	ZNF185
C1QBP	EIF3J	HNRNPA2B1	NCOA6	RBM10	TACC1	ZYX
CALR	EPB41	HNRNPM	NDUFA7	RBM27	TAGLN	
CATD	EXOG	HRSP12	NDUFB11	RPL19	TFAM	
CD3EAP	FAM136A	IRF2BP2	NES	SBSN	THBS1	
CDC5L	FAM3C	ISCA2	NIPSNAP3A	SCAF11	TJP1	
CDH13	FIS1	ITGA6	P32004	SCAMP1	TNS1	

B: Proteins higher in control siRNA-treated samples

AAK1	CFL2	F8W031	H0YHG0	P14174	PPM1G	S100A2	TIMM8A
ADNP	CHCHD3	FAM192A	HSPB8	PAGE2B	PSAP	SDC4	TSNAX
ARPP19	CNN3	FAM195B	JAM3	PALLD	PSMA5	SKP1	UBAP2
ATOX1	DENR	G3BP1	LDHA	PCBP1	PSMA7	SNAP23	UBE2V2
CAPN2	DLGAP5	GLO1	MAP4	PDLIM7	PSMB3	SPAG7	UBXN1
CEP131	ECI1	GPHN	MYL12A	PGM2	PSME3	SUPT5H	UBXN7
CFDP1	EEF1B2	GSTP1	NPC2	PLP2	PVRL3	TIA1	

* as explained in text, some gene names have been replaced with Uniprot identifiers.

6.6.5 GO term overrepresentation analysis 1: Proteins that were more abundant in si*UNR*-treated U2OS cells than in equivalent control siRNA-treated U2OS cells

There were seven over represented biological process GO terms using the proteins up-regulated following si*UNR* treatment (Table 6.25A). There were nine overrepresented molecular function GO terms (Table 6.25B). There were 34 overrepresented cellular component GO terms, the top ten of which are presented by p-value (Table 6.25C) and by fold enrichment (Table 6.25D).

Table 6.25: Top overrepresented GO terms associated with proteins that were higher in abundance in unstressed si*UNR*-treated U2OS cells than in unstressed control siRNA-treated U2OS cells, by p-value or fold enrichment (as stated)

A: Biological process GO terms, by p-value

GO biological process complete	Ref	Obs	Exp	FE	p-value
mRNA processing (GO:0006397)	445	14	2.12	6.6	2.48E-04
RNA splicing (GO:0008380)	377	12	1.8	6.68	2.29E-03
mRNA metabolic process (GO:0016071)	627	15	2.99	5.02	2.59E-03
biological_process (GO:0008150)	17008	97	81.1	1.2	1.55E-02
cellular component organization or biogenesis (GO:0071840)	5481	48	26.13	1.84	1.87E-02
cellular component organization (GO:0016043)	5318	47	25.36	1.85	2.02E-02
cellular process (GO:0009987)	14523	89	69.25	1.29	2.33E-02

B: Molecular function GO terms, by p-value

GO molecular function complete	Ref	Obs	Exp	FE	p-value
poly(A) RNA binding (GO:0044822)	1168	31	5.57	5.57	5.11E-12
RNA binding (GO:0003723)	1631	34	7.78	4.37	1.70E-10
protein binding (GO:0005515)	10901	77	51.98	1.48	5.95E-04
binding (GO:0005488)	14421	90	68.76	1.31	1.22E-03
nucleic acid binding (GO:0003676)	4068	41	19.4	2.11	1.50E-03
cell adhesion molecule binding (GO:0050839)	456	12	2.17	5.52	5.30E-03
molecular_function (GO:0003674)	17134	97	81.7	1.19	9.27E-03
cadherin binding (GO:0045296)	295	9	1.41	6.4	3.45E-02
laminin binding (GO:0043236)	31	4	0.15	27.06	4.45E-02

C: Cellular component GO terms, by p-value

GO cellular component complete	Ref	Obs	Exp	FE	p-value
intracellular organelle part (GO:0044446)	8167	74	38.94	1.9	1.52E-09
organelle part (GO:0044422)	8351	74	39.82	1.86	5.50E-09
intracellular part (GO:0044424)	13819	93	65.89	1.41	1.83E-07
organelle (GO:0043226)	13070	90	62.32	1.44	5.16E-07
adherens junction (GO:0005912)	684	19	3.26	5.83	7.66E-07
intracellular organelle (GO:0043229)	12049	86	57.45	1.5	9.23E-07
cytoplasm (GO:0005737)	10781	81	51.41	1.58	9.76E-07
anchoring junction (GO:0070161)	702	19	3.35	5.68	1.17E-06
intracellular (GO:0005622)	14192	93	67.67	1.37	1.52E-06
intracellular organelle lumen (GO:0070013)	4379	48	20.88	2.3	1.90E-06

D: Cellular component GO terms, by fold enrichment

GO cellular component complete	Ref	Obs	Exp	FE	p-value
spliceosomal complex (GO:0005681)	178	7	0.85	8.25	3.29E-02
focal adhesion (GO:0005925)	393	14	1.87	7.47	8.28E-06
cell-substrate adherens junction (GO:0005924)	396	14	1.89	7.41	9.11E-06
cell-substrate junction (GO:0030055)	400	14	1.91	7.34	1.03E-05
cell-cell adherens junction (GO:0005913)	320	9	1.53	5.9	3.16E-02
adherens junction (GO:0005912)	684	19	3.26	5.83	7.66E-07
anchoring junction (GO:0070161)	702	19	3.35	5.68	1.17E-06
mitochondrial part (GO:0044429)	998	19	4.76	3.99	3.06E-04
cell junction (GO:0030054)	1371	22	6.54	3.37	5.50E-04
organelle envelope (GO:0031967)	1145	18	5.46	3.3	9.75E-03

6.6.6 Consideration of the overrepresented GO terms from proteins that were more abundant in si*UNR*-treated U2OS cells than in equivalent control siRNA-treated U2OS cells

As was observed with the unstressed HeLa samples, the strongest p-value for the overrepresented biological process GO terms ($p=2.48x10^{-4}$) are higher than that for the other two groups of GO terms. Unlike the situation with the HeLa samples, however, the strongest molecular function GO term ($p=5.11x10^{-12}$) was stronger than that for the best cellular component GO term ($p=1.52x10^{-9}$) (Table 6.25).

The top three overrepresented biological process GO terms were related to RNA processing, lending further support to the idea that UNR may affect protein expression at multiple levels (Table 6.25A). The strongest molecular function GO term by p-value, 'poly(A) RNA binding', had the lowest p-value of all overrepresented GO terms and the second strongest p-value was the more general 'RNA binding' (Table 6.25B). Interestingly, there were only 3 additional observations for the more general GO term, thereby implying that UNR is involved in regulating the expression of poly(A) binding proteins in a fairly specific way. In addition to this, the third strongest p value for a molecular function GO term was around 3,500,000 times higher than that for the second (i.e. $1.16x10^{8}$ higher than that for 'poly(A) RNA binding') – further implying that UNR may have a specific role in regulating poly(A) binding. The strongest p-value for an overrepresented non-general GO term (taken to be a GO term with over 1000 annotated proteins) was 'adherens junction' with a Bonferroni adjusted p-value of $7.66x10^{-7}$ (Table 6.25C). This GO term had the highest fold change of any of the top 10 overrepresented cellular component GO terms by p-value and also made the top ten overrepresented cellular component GO terms by fold change (Table 6.25D).

These observations lead to a hypothesis whereby UNR may have a general role in suppressing the expression of a wide array of proteins but that it could have a specific role in suppressing the expression of proteins involved in poly(A) binding and those that localise to adherens junctions.

6.6.7 GO term overrepresentation analysis 2: Proteins that were less abundant in si*UNR*-treated U2OS cells than in equivalent control siRNA-treated U2OS cells

There were eight overrepresented biological process GO terms using the proteins down-regulated following si*UNR* treatment (Table 6.26A). There were two overrepresented molecular function GO terms (Table 6.26B). There were 14 overrepresented cellular component GO terms, the top ten of which are presented by p-value (Table 6.26C) and by fold enrichment (Table 6.26D).

Table 6.26: Top overrepresented GO terms associated with proteins that were higher in abundance in unstressed control siRNA-treated U2OS cells than in unstressed si*UNR*-treated U2OS cells, by p-value or fold enrichment (as stated)

A: Biological process GO terms, by p-value

GO biological process complete	Ref	Obs	Exp	FE	p-value
positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition (GO:0051437)	76	5	0.2	25.09	1.55E-02
regulation of proteolysis involved in cellular protein catabolic process (GO:1903050)	228	7	0.6	11.71	1.91E-02
positive regulation of ubiquitin protein ligase activity (GO:1904668)	81	5	0.21	23.54	2.11E-02
NIK/NF-kappaB signaling (GO:0038061)	85	5	0.22	22.43	2.67E-02
regulation of ubiquitin protein ligase activity (GO:1904666)	85	5	0.22	22.43	2.67E-02
regulation of cellular protein catabolic process (GO:1903362)	250	7	0.66	10.68	3.49E-02
positive regulation of protein ubiquitination involved in ubiquitin-dependent protein catabolic process (GO:2000060)	92	5	0.24	20.72	3.91E-02
positive regulation of proteolysis involved in cellular protein catabolic process (GO:1903052)	170	6	0.45	13.46	4.85E-02

B: Molecular function GO terms, by p-value

GO molecular function complete	Ref	Obs	Exp	FE	p-value
protein binding (GO:0005515)	10901	50	28.59	1.75	1.63E-06
binding (GO:0005488)	14421	52	37.82	1.37	8.43E-03

C: Cellular component GO terms, by p-value

GO cellular component complete	Ref	Obs	Exp	FE	p-value
anchoring junction (GO:0070161)	702	12	1.84	6.52	2.92E-04
extracellular exosome (GO:0070062)	2735	22	7.17	3.07	7.29E-04
extracellular vesicle (GO:1903561)	2749	22	7.21	3.05	7.96E-04
extracellular organelle (GO:0043230)	2750	22	7.21	3.05	8.01E-04
proteasome complex (GO:0000502)	69	5	0.18	27.63	1.51E-03
endopeptidase complex (GO:1905369)	69	5	0.18	27.63	1.51E-03
adherens junction (GO:0005912)	684	11	1.79	6.13	1.82E-03
cytoplasm (GO:0005737)	10781	45	28.27	1.59	3.53E-03
membrane-bounded vesicle (GO:0031988)	3547	24	9.3	2.58	4.10E-03
extracellular region part (GO:0044421)	3819	25	10.02	2.5	4.12E-03

D: Cellular component GO terms, by fold enrichment

GO cellular component complete	Ref	Obs	Exp	FE	p-value
proteasome core complex (GO:0005839)	22	3	0.06	52	3.75E-02
proteasome complex (GO:0000502)	69	5	0.18	27.63	1.51E-03
endopeptidase complex (GO:1905369)	69	5	0.18	27.63	1.51E-03
peptidase complex (GO:1905368)	90	5	0.24	21.18	5.47E-03
anchoring junction (GO:0070161)	702	12	1.84	6.52	2.92E-04
adherens junction (GO:0005912)	684	11	1.79	6.13	1.82E-03
extracellular exosome (GO:0070062)	2735	22	7.17	3.07	7.29E-04
extracellular vesicle (GO:1903561)	2749	22	7.21	3.05	7.96E-04
extracellular organelle (GO:0043230)	2750	22	7.21	3.05	8.01E-04
membrane-bounded vesicle (GO:0031988)	3547	24	9.3	2.58	4.10E-03

6.6.8 Consideration of the overrepresented GO terms from proteins that were less abundant in si*UNR*-treated U2OS cells than in equivalent control siRNA-treated U2OS cells

In terms of the data entered into the GO tool, it was noted that there were fewer proteins that had their expression lowered by si*UNR* treatment (55) than had it increased (100) in unstressed U2OS cells (section 6.6.4). In HeLa, the numbers were 127 (increased on si*UNR*) and 104 (decreased on si*UNR*). It should be noted that different programs were used to analyse the two sets of data and that different p-value cut-offs were used. In both HeLa and U2OS, the strongest p-value for their overrepresented biological process and molecular function GO terms were higher for proteins that decreased on si*UNR* treatment relative to those that increased. With the U2OS samples, there was also a very substantial weakening of the best p-value for overrepresented cellular component GO terms on moving from proteins that increased on si*UNR* treatment to those than decreased (Tables 6.25C, 6.26C). It should be remembered that there were approximately twice as many proteins considered significantly up-regulated by si*UNR* treatment in unstressed U2OS cells (100) than proteins significantly down-regulated by si*UNR* treatment (55) fed into the GO tool.

There were no strong p-values for any of the overrepresented biological process GO terms generated using the proteins that decreased in abundance in U2OS cells following treatment with si*UNR* (Table 6.26A). All of the significantly overrepresented biological process GO terms had strong fold enrichments of over 10 (Table 6.26A). This was to be expected as there were fewer proteins to generate a significant p-value with a small fold enrichment. Whereas the top

hits for HeLa were generally related to oxidative phosphorylation, the top hits in unstressed U2OS cells were more related to ubiquitin-related proteolysis (Table 6.26A). There were only two significantly overrepresented molecular function GO terms, and those were both very general (Table 6.26B). The strongest p-values among the overrepresented cellular component GO terms included 'anchoring junction', the related 'adherens junction' and a set of three related GO terms including 'extracellular exosome'. The list also included 'proteasome complex', tying in with the ubiquitin-related biological process GO terms (Table 6.26C). There were four related cellular component GO terms with large fold enrichments over 21; these were related to the proteasome/proteolysis (Table 6.26D). It would appear, therefore, that UNR both promotes and suppresses the expression of proteins that localise to adherens junctions.

6.7 Analysis of data obtained from U2OS cells that were stressed with sodium arsenite following previous treatment with si*UNR* or a control siRNA

As Progenesis was considered more powerful than Scaffold for protein quantification (see chapter 4), it was decided to use Progenesis to analyse the mass spectrometry data obtained using arsenite-stressed U2OS cell lysates. This was considered more useful than having the stressed and unstressed samples analysed with the same program. The result of that decision is that the analysis of the stressed U2OS sample would be considered better and direct comparison between findings between the stressed and unstressed samples should be treated with caution.

Automatic processing of the stressed U2OS lysate samples showed that there were similar numbers of MS peaks and MS/MS counts across the six samples, as well as similar total ion intensities (Table 6.27).

Table 6.27: Data from Progenesis pertaining to stressed U2OS knockdown samples

Repeat	knockdown	MS peak count ($x10^7$)	MS/MS count	total ion intensity ($x10^{10}$)
Day 1	si*UNR*	1.75	40529	1.48
Day 1	control	1.75	44478	2.04
Day 2	si*UNR*	1.76	39195	1.42
Day 2	control	1.76	42191	1.84
Day 3	si*UNR*	1.87	34883	1.44
Day 3	control	1.79	43530	2.10

The samples were processed as for the HeLa knockdown samples (see section 6.3.3) except that, features coming off before 25 minutes (as dirty) or after 150 minutes (as the U2OS samples were run for 3 hours and not 4 hours) were rejected.

As with the HeLa knockdown samples, PCA analysis was carried out with and without the removal of proteins with ANOVA p-values over 0.1 (Figure 6.8).

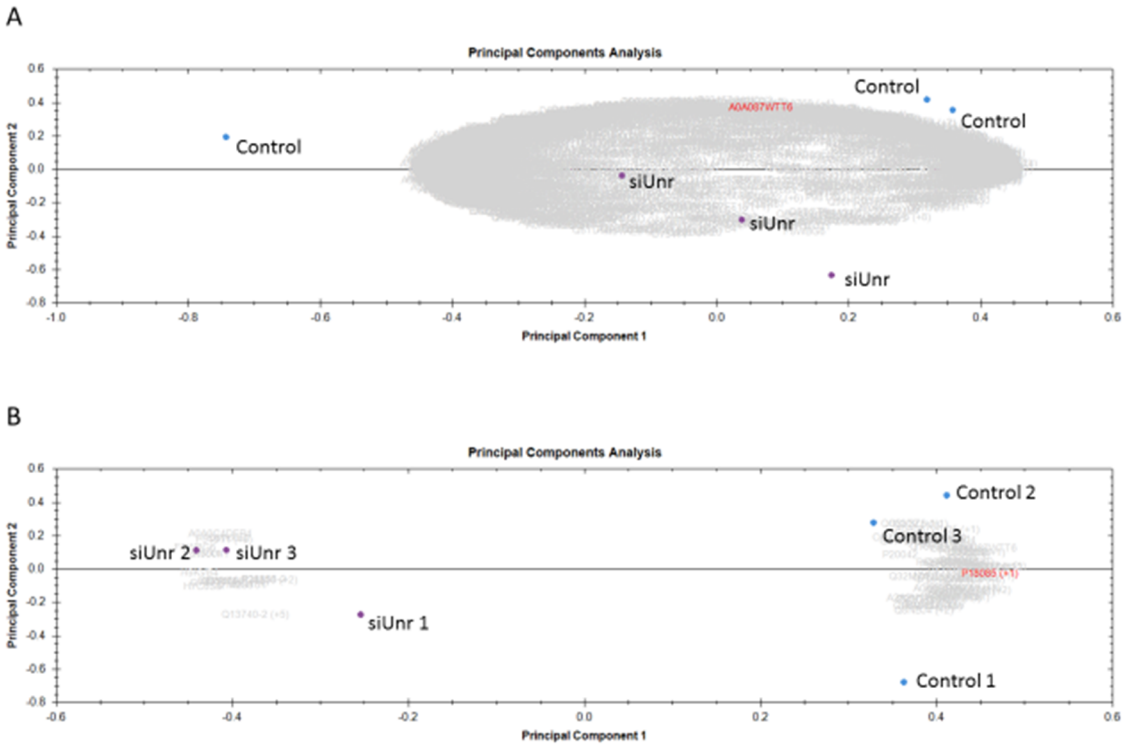


Figure 6.8: Progenesis-generated PCA plots for HeLa knockdown data. Proteins with all ANOVA p-values (A) or only those with ANOVA p-values under 0.1 (B) were included. See text for further information how the data were processed. The numbers in (B) are associated with the days upon which the lysates were made (i.e. the repeats). Purple circles = si*UNR* samples (siUnr); blue circles = control siRNA. In terms of variance accounted for – A: unrecorded, B: (PC1 = 77.46%, PC2 = 8.17%). The grey writing in the background makes up the protein score plot (see section 4.9.2). The red writing refers to an automatically highlighted protein ID.

The PCA analysis clearly showed that the removal of the proteins with high ANOVA values was beneficial as doing so resulted in PC1, accounting for 77.46% of the variance, clearly delineating the control samples from the si*UNR* samples. When all proteins were included, the siRNA

treatment was not the main source of variation in the data (Figure 6.8). It was therefore decided to use only those proteins with ANOVA p-values less than 0.1.

Whilst there was only one UNR peptide detected across all samples, its distribution provided additional support to the idea that the knockdowns had been successful (Figure 6.9).

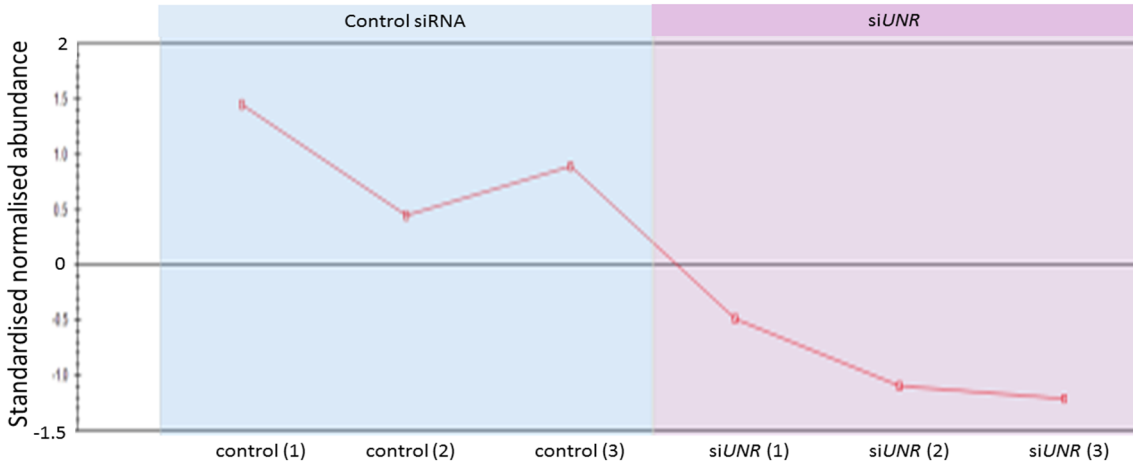


Figure 6.9: Standardised normalised abundances for a peptide assigned to UNR from the arsenite stressed U2OS knockdown samples. The repeats in the blue column are the control siRNA-treated samples and the repeats in the purple column are the si*UNR*-treated samples.

6.7.1 Detection of proteins that are differentially expressed in U2OS cells stressed with sodium arsenite following previous treatment with either si*UNR* or a control siRNA

As with the HeLa knockdown experiment, the protein abundance data was then exported to Microsoft Excel and paired t-tests were used to find proteins that potentially have their expression modulated by UNR. Ratios of the sum of the si*UNR*-treated sample abundances for each protein to the sum of the sample abundances for the control siRNA-treated sample abundances were also calculated. From these values, putative UNR-mediated differentially expressed proteins were selected as those with:

1) t-test p-value less than 0.1, higher in si*UNR* samples

2) t-test p-value less than 0.1, higher in control siRNA samples

3) si*UNR*/control siRNA ratio greater than 10

4) control siRNA/si*UNR* ratio greater than 10

6.7.2 Discovery of proteins with significantly different expression levels in U2OS cells stressed with sodium arsenite following previous treatment with either si*UNR* or a control siRNA

The changes in protein levels were explored in turn for each of the four groups stated in previous sections.

1) t-test p-value less than 0.1, higher in si*UNR* samples

In total, there were 13 proteins that had p-values under 0.05 and higher abundances in the si*UNR*-treated samples. As this was a very low number compared to the equivalent value for HeLa, it was decided to extend the p-value cut-off to p<0.1. That gave an additional 6 hits, giving 19 in total. The top ten hits that were higher in the si*UNR*-treated samples, by t-test p-value, are presented in Table 6.28.

Table 6.28: Top ten proteins whose expression is higher in arsenite stressed si*UNR*-treated U2OS cells than in arsenite stressed control siRNA-treated U2OS cells, by t-test p-value

Protein	p-value
Peptide-N(4)-(N-acetyl-beta-glucosaminyl)asparagine amidase (Fragment) GN=NGLY1	0.000493
Carbohydrate sulfotransferase 2 GN=CHST2	0.003760
Uncharacterized protein (Fragment)	0.006895
Isoform 2 of CD166 antigen GN=ALCAM	0.008512
Peroxiredoxin-6 GN=PRDX6	0.010145
Isoform 5 of PDZ and LIM domain protein 2 GN=PDLIM2	0.015700
Protein ELFN1 GN=ELFN1	0.018046
Trinucleotide repeat-containing gene 18 protein GN=TNRC18	0.019842
Paralemmin-1 GN=PALM	0.036128
Caskin-2 GN=CASKIN2	0.037437

2) t-test p-value less than 0.1, higher in control siRNA samples

In total, there were 68 proteins that had p-values under 0.1 and higher abundances in the control siRNA-treated samples. The top ten hits that were higher in the control siRNA-treated samples, by t-test p-value, are presented in Table 6.29. This shows that, unlike the situation with unstressed HeLa and U2OS cells, there were more proteins down-regulated rather than up-regulated following *UNR* knockdown in U2OS cells.

Table 6.29: Top ten proteins whose expression is higher in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells, by t-test p-value

Description	p-value
Caprin-1 GN=CAPRIN1	0.002978
Inorganic pyrophosphatase GN=PPA1	0.006884
Cold shock domain-containing protein E1 GN=CSDE1	0.007902
Syntaxin-12 GN=STX12	0.008339
Signal recognition particle 14 kDa protein GN=SRP14	0.009660
Glutathione S-transferase omega-1 (Fragment) GN=GSTO1	0.010657
Isoform 2 of Protein enabled homolog GN=ENAH	0.011367
Pumilio homolog 2 GN=PUM2	0.013507
PDZ and LIM domain protein 5 GN=PDLIM5	0.014547
DNA fragmentation factor subunit alpha GN=DFFA	0.014795

3) si*UNR*/control siRNA ratio greater than 10

No proteins were considered significant on the grounds of having si*UNR*/control siRNA abundance ratios above 10:1. The highest ratio was 1.8:1 for Trinucleotide repeat-containing gene 18 protein (GN=TNRC18).

4) control siRNA/si*UNR* ratio greater than 10

18 proteins were considered significant on the grounds of having control siRNA/si*UNR* abundance ratios above 10:1. These were largely ribosomal proteins. The top ten by ratio are presented in Table 6.30. It was noted that proteins that are down-regulated following si*UNR* treatment are down-regulated by a greater degree than those that are up-regulated.

Table 6.30: Top ten proteins whose expression is higher in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells, by fold enrichment

Description	p-value	control siRNA/si*UNR*
40S ribosomal protein S29 GN=RPS29	0.2968	4690811
Alanine--tRNA ligase, cytoplasmic GN=AARS	0.3304	62.99
60S ribosomal protein L27 GN=RPL27	0.3440	57.62
40S ribosomal protein S13 GN=RPS13	0.3146	31.05
60S ribosomal protein L10 GN=RPL10	0.3560	28.20
60S ribosomal protein L27a GN=RPL27A	0.3450	25.02
60S ribosomal protein L34 GN=RPL34	0.3478	23.33
60S ribosomal protein L8 GN=RPL8	0.3354	21.00
40S ribosomal protein S17 GN=RPS17	0.3180	19.45
60S ribosomal protein L18a GN=RPL18A	0.3568	19.06

6.8 GO-term overrepresentation analyses using proteins differentially expressed in sodium arsenite-treated U2OS cells that had previously been treated with either si*UNR* or control siRNA

In order to carry out GO term analyses on the arsenite stressed U2OS knockdown data, the proteins that were statistically up-regulated by si*UNR* treatment were used alone whereas the down-regulated proteins were merged with those that had were present in the control siRNA sample at a level at least ten times greater than that in the si*UNR* samples.

6.8.1 GO-term overrepresentation analysis on proteins that were more abundant in si*UNR*-treated U2OS cells, that were subsequently treated with sodium arsenite, than in equivalent control siRNA-treated U2OS cells

The 19 proteins that were significantly more abundant in the si*UNR*-treated samples were fed into the GO tool at http://amigo.geneontology.org/rte which recognised all of them. The list, as used, is presented in Table 6.31.

Table 6.31: Gene names of proteins higher in abundance in arsenite stressed si*UNR*-treated U2OS cells than in arsenite stressed control siRNA-treated U2OS cells

ALCAM	CHST2	H7C0S8	PALM	RAPGEF3
BBX	CNN1	KNSTRN	PDLIM2	TM4SF1
CALD1	DNMT1	NGLY1	PRDX6	TNRC18
CASKIN2	ELFN1	NLRP10	PTK2B	

There were no enriched Biological Process or Molecular Function GO terms using proteins up-regulated following si*UNR* treatment in arsenite stressed U2OS cells. There were four significantly enriched cellular component GO terms (Table 6.32).

Table 6.32: Overrepresented cellular component GO terms associated with proteins that were higher in abundance in arsenite stressed si*UNR*-treated U2OS cells than in arsenite stressed control siRNA-treated U2OS cells, by p-value

GO cellular component complete	Ref	Obs	Exp	FE	p-value
adherens junction (GO:0005912)	684	6	0.62	9.68	2.92E-02
anchoring junction (GO:0070161)	702	6	0.64	9.43	3.38E-02
cortical actin cytoskeleton (GO:0030864)	65	3	0.06	50.94	3.59E-02
somatodendritic compartment (GO:0036477)	720	6	0.65	9.2	3.90E-02

6.8.2 Consideration of GO term analysis on proteins that were more abundant in si*UNR*-treated U2OS cells, that were subsequently treated with sodium arsenite, than in equivalent control siRNA-treated U2OS cells

The low number of proteins that significantly increased in abundance following si*UNR* treatment was presumably responsible for the lack of Bonferroni-corrected significant biological process or molecular function GO terms. It was interesting that ‘adherens junction’ was the top cellular component hit. It would appear that UNR is strongly associated with the expression levels of proteins found at adherens junctions.

6.8.3 GO-term overrepresentation analysis on proteins that were less abundant in si*UNR*-treated U2OS cells, that were subsequently treated with sodium arsenite, than in equivalent control siRNA-treated U2OS cells

The proteins considered significantly down-regulated by p-value and fold change following si*UNR* treatment were listed by gene name and any duplicate values were removed. These were then initially fed into the GO tool at http://amigo.geneontology.org/rte. That returned two genes with multiple entries. Acceptable Uniprot identifiers were obtained for those genes and were used to replace them. The gene/protein list was re-entered into the GO tool. One protein retained multiple entries, this time for two genes. As the two genes encoded the same protein, the protein was retained in the list. That 86 entry list, as used, is presented in Table 6.33.

Table 6.33: Gene/protein list corresponding to proteins down-regulated on si*UNR* treatment

AARS	CLIC1	KPNA2	PPP1R13L	RPL36AL	TJP2
ACTR2	CSDE1	LAP3	PRRC2A	RPL8	TK1
AIMP1	CYR61	LETM2	PTPN12	RPP30	TLK1
ANXA5	DFFA	LIMD1	PUM2	RPS13	TUBA1B
ARF4	DHRS2	LRRFIP1	Q13748	RPS17	TUBA1C
ARPC1B	DYNLL2	MCCC2	Q99575	RPS27	TUBB2A
ATP6V1G1	EEF1D	MED15	RAB29	RPS29	TUBB4B
BOD1L1	EFTUD2	NAP1L1	RPL10	RPS4X	TUBB6
C11orf98	EIF2S2	NCOA3	RPL13A	RPSA	UBE2A
C1orf122	ENAH	NLRP8	RPL18A	SEPHS2	YWHAB
CAPRIN1	ERH	NOP56	RPL27	SRP14	YWHAZ
CAPZB	FAM103A1	NPM3	RPL27A	STX12	
CDK14	GOLM1	NUMB	RPL32	SUPT5H	
CDKN2AIPNL	GSTO1	PDLIM5	RPL34	TAF7	
CKAP2L	IARS2	PPA1	RPL36A	TCERG1	

1) Enriched Biological Process GO terms using proteins down-regulated following si*UNR* treatment in U2OS cells that were subsequently treated with sodium arsenite

There were 85 significantly overrepresented biological process GO terms among the proteins down-regulated by si*UNR* treatment. The top ten over-represented GO terms by p-value are presented in Table 6.34 and the top ten by fold enrichment in Table 6.35.

Table 6.34: Top ten overrepresented biological process GO terms associated with proteins that were higher in abundance in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells, by p-value

GO biological process complete	Ref	Obs	Exp	FE	p-value
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	95	16	0.39	40.6	2.28E-17
protein targeting to ER (GO:0045047)	101	16	0.42	38.19	5.96E-17
cotranslational protein targeting to membrane (GO:0006613)	103	16	0.43	37.45	8.11E-17
establishment of protein localization to endoplasmic reticulum (GO:0072599)	105	16	0.44	36.73	1.10E-16
translation (GO:0006412)	455	24	1.89	12.72	4.47E-16
multi-organism metabolic process (GO:0044033)	145	17	0.6	28.26	4.89E-16
peptide biosynthetic process (GO:0043043)	479	24	1.99	12.08	1.43E-15
protein localization to endoplasmic reticulum (GO:0070972)	125	16	0.52	30.86	1.67E-15
peptide metabolic process (GO:0006518)	625	26	2.59	10.03	3.10E-15
viral transcription (GO:0019083)	115	15	0.48	31.44	1.84E-14

Table 6.35: Top ten overrepresented biological process GO terms associated with proteins that were higher in abundance in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells, by fold enrichment

GO biological process complete	Ref	Obs	Exp	FE	p-value
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	95	16	0.39	40.6	2.28E-17
protein targeting to ER (GO:0045047)	101	16	0.42	38.19	5.96E-17
cotranslational protein targeting to membrane (GO:0006613)	103	16	0.43	37.45	8.11E-17
establishment of protein localization to endoplasmic reticulum (GO:0072599)	105	16	0.44	36.73	1.10E-16
viral transcription (GO:0019083)	115	15	0.48	31.44	1.84E-14
protein localization to endoplasmic reticulum (GO:0070972)	125	16	0.52	30.86	1.67E-15
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	119	15	0.49	30.39	3.03E-14
viral gene expression (GO:0019080)	126	15	0.52	28.7	6.98E-14
multi-organism metabolic process (GO:0044033)	145	17	0.6	28.26	4.89E-16
translational initiation (GO:0006413)	154	16	0.64	25.04	4.29E-14
protein targeting to membrane (GO:0006612)	164	16	0.68	23.52	1.14E-13

2) Enriched Molecular Function GO terms using proteins down-regulated following si*UNR* treatment in U2OS cells that were subsequently treated with sodium arsenite

There were 16 significantly overrepresented molecular function GO terms among the proteins down-regulated by si*UNR* treatment. The top ten over-represented GO terms by p-value are presented in Table 6.36 and the top ten by fold enrichment in Table 6.37.

Table 6.36: Top ten overrepresented molecular function GO terms associated with proteins that were higher in abundance in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells, by p-value

GO molecular function complete	Ref	Obs	Exp	FE	p-value
RNA binding (GO:0003723)	1631	39	6.77	5.76	2.79E-17
poly(A) RNA binding (GO:0044822)	1168	31	4.85	6.4	5.69E-14
structural constituent of ribosome (GO:0003735)	226	16	0.94	17.07	5.06E-12
structural molecule activity (GO:0005198)	784	24	3.25	7.38	2.54E-11
heterocyclic compound binding (GO:1901363)	5953	55	24.7	2.23	4.12E-08
organic cyclic compound binding (GO:0097159)	6037	55	25.04	2.2	7.49E-08
cadherin binding involved in cell-cell adhesion (GO:0098641)	277	12	1.15	10.44	5.32E-06
protein binding involved in cell-cell adhesion (GO:0098632)	287	12	1.19	10.08	7.87E-06
protein binding involved in cell adhesion (GO:0098631)	292	12	1.21	9.91	9.53E-06
cadherin binding (GO:0045296)	295	12	1.22	9.81	1.07E-05

Table 6.37: Top ten overrepresented molecular function GO terms associated with proteins that were higher in abundance in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells, by fold enrichment

GO molecular function complete	Ref	Obs	Exp	FE	p-value
structural constituent of cytoskeleton (GO:0005200)	109	8	0.45	17.69	5.75E-05
structural constituent of ribosome (GO:0003735)	226	16	0.94	17.07	5.06E-12
cadherin binding involved in cell-cell adhesion (GO:0098641)	277	12	1.15	10.44	5.32E-06
protein binding involved in cell-cell adhesion (GO:0098632)	287	12	1.19	10.08	7.87E-06
protein binding involved in cell adhesion (GO:0098631)	292	12	1.21	9.91	9.53E-06
cadherin binding (GO:0045296)	295	12	1.22	9.81	1.07E-05
GTPase activity (GO:0003924)	233	9	0.97	9.31	1.61E-03
structural molecule activity (GO:0005198)	784	24	3.25	7.38	2.54E-11
cell adhesion molecule binding (GO:0050839)	456	13	1.89	6.87	1.47E-04
poly(A) RNA binding (GO:0044822)	1168	31	4.85	6.4	5.69E-14

3) Enriched Cellular component GO terms using proteins down-regulated following si*UNR* treatment in U2OS cells that were subsequently treated with sodium arsenite

There were 46 significantly overrepresented cellular component GO terms among the proteins down-regulated by si*UNR* treatment. The top ten over-represented GO terms by p-value are presented in Table 6.38 and the top ten by fold enrichment in Table 6.39.

Table 6.38: Top ten overrepresented cellular component GO terms associated with proteins that were higher in abundance in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells, by p-value

GO cellular component complete	Ref	Obs	Exp	FE	p-value
cytosolic ribosome (GO:0022626)	122	16	0.51	31.61	1.78E-16
adherens junction (GO:0005912)	684	26	2.84	9.16	4.25E-15
anchoring junction (GO:0070161)	702	26	2.91	8.93	7.93E-15
ribosomal subunit (GO:0044391)	174	16	0.72	22.17	4.42E-14
cytosolic part (GO:0044445)	243	17	1.01	16.86	3.62E-13
cytosol (GO:0005829)	3487	48	14.47	3.32	4.21E-13
intracellular ribonucleoprotein complex (GO:0030529)	775	24	3.22	7.47	9.60E-12
ribonucleoprotein complex (GO:1990904)	776	24	3.22	7.46	9.87E-12
ribosome (GO:0005840)	252	16	1.05	15.31	1.29E-11
intracellular non-membrane-bounded organelle (GO:0043232)	3990	47	16.55	2.84	4.64E-10

Table 6.39: Top ten overrepresented cellular component GO terms associated with proteins that were higher in abundance in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells, by fold enrichment

GO cellular component complete	Ref	Obs	Exp	FE	p-value
cytosolic large ribosomal subunit (GO:0022625)	68	10	0.28	35.45	5.29E-10
cytosolic small ribosomal subunit (GO:0022627)	45	6	0.19	32.14	5.48E-05
cytosolic ribosome (GO:0022626)	122	16	0.51	31.61	1.78E-16
large ribosomal subunit (GO:0015934)	104	10	0.43	23.18	3.28E-08
ribosomal subunit (GO:0044391)	174	16	0.72	22.17	4.42E-14
small ribosomal subunit (GO:0015935)	70	6	0.29	20.66	7.15E-04
cytosolic part (GO:0044445)	243	17	1.01	16.86	3.62E-13
ribosome (GO:0005840)	252	16	1.05	15.31	1.29E-11
focal adhesion (GO:0005925)	393	17	1.63	10.43	7.95E-10
cell-substrate adherens junction (GO:0005924)	396	17	1.64	10.35	8.96E-10

6.8.4 Consideration of GO term analysis on proteins that were less abundant in si*UNR*-treated U2OS cells, that were subsequently treated with sodium arsenite, than in equivalent control siRNA-treated U2OS cells

The first observation regarding the overrepresented GO terms was that, whilst the number of significant GO terms in each group was quite different (biological process = 85, molecular

function = 16, cellular component = 46), the strongest p-value from each group was quite similar (biological process = $2.28x10^{-17}$, molecular function = $2.79x10^{-17}$ with cellular component being slightly higher = $1.78x10^{-16}$).

In terms of the most significantly overrepresented biological process GO terms, an initial observation was that one of the top ten by p-value (Table 6.34) and two of the top ten by fold enrichment (Table 6.35) were to do with viral gene expression. This is in contrast to the unstressed HeLa samples that had significant virus-related GO terms located using proteins that were more abundant in si*UNR*-treated samples. A group of proteins generated four closely related significant biological process GO terms that implied that *UNR* knockdown resulted in a reduction of expression of proteins involved in signal recognition particle (SRP)-dependent cotranslational protein targeting to the endoplasmic reticulum. This would imply that UNR would normally increase or maintain the expression of those proteins in the presence of arsenite stress. Whilst the composition changes between eukaryotes, eubacteria and archaea, the general ribonucleoprotein SRP complex is ubiquitous in nature (Luirink & Sinning 2004). Given the conserved nature of the SRP complex, it would be interesting to see if this observation could be repeated in other animals, such as *Drosophila melanogaster*. The next most significantly overrepresented GO term by p-value was 'translation'. The reduction of proteins involved in translation on arsenite treatment of si*UNR*-treated cells relative to arsenite treated control siRNA-treated cells suggests that UNR could increase or maintain translation under arsenite stress. This would offer a mechanism by which UNR could mediate the cellular response to oxidative stress. The top two overrepresented molecular function GO terms by p-value among proteins up-regulated by si*UNR* treatment in unstressed U2OS cells were 'poly(A) RNA binding' and 'RNA binding' (section 6.6.6). This was not observed when the cells were stressed prior to harvesting (section 6.8.1). Interestingly, the same two GO terms were the top two among proteins down-regulated in si*UNR*-treated U2OS cells that were stressed prior to harvesting (Table 6.36). Indeed, the p-values were even stronger. Another top hit by p-value (Table 6.36) and fold enrichment (Table 6.35) was 'structural component of the ribosome'. Ribosome-related GO terms were among the most significant overrepresented cellular component GO terms, both by p-value (Table 6.38) and fold enrichment (Table 6.39). Indeed, the strongest p-value was for 'cytosolic ribosome'. The second strongest p-value ($p = 4.25x10^{-15}$) was for 'adherens junction', reinforcing the idea that UNR is intimately involved in regulation of protein levels at adherens junctions (Table 6.38).

6.9 Cellular Component GO-term overrepresentation analysis revisited for proteins that were less abundant in si*UNR*-treated U2OS cells, that were subsequently treated with sodium arsenite, than in equivalent control siRNA-treated U2OS cells

It seemed clear that the list of proteins down-regulated on si*UNR* treatment was likely to contain a lot of true positives. This assumption was based on the GO term analysis; if the proteins had largely been random false positives, it would be extremely unlikely that they would generate such strong p-values. This point was particularly important as the analysis as a whole was based upon the GO term analysis being a safety net against the acceptance of a large number of false positives as multiple testing correction was not possible at the earlier significance-detection stage of the analysis.

As a test of the assumption that the significance detection level was adequately stringent, accepting more hits by loosening the stringency should strengthen the overrepresented GO-term p-values. Should the overrepresented GO-term p-values weaken upon the addition of the additional proteins, it would imply that the p-value cut-offs had been too liberal and that would be less positive in terms of confidence in the overall analysis.

To do this, the protein universe that was used to detect significant proteins based on their t-test p-values and abundance ratios was expanded to include proteins with any ANOVA value. If this method produced more putative significant proteins and also stronger p-values for specific overrepresented GO terms, with particular reference to the recurrent 'adherens junction' cellular component GO term, that would strongly imply that most hits were likely to be true positives and that more true positives were excluded by the original detection criteria.

To avoid repeating the entire analysis, the new protein list is presented below (Table 6.40) and then used exclusively for cellular component overrepresentation GO term analysis (Table 6.41).

Table 6.40: Gene list corresponding to proteins that are higher in arsenite stressed control siRNA-treated U2OS cells over arsenite stressed si*UNR*-treated U2OS cells

AARS	CDK14	FUS	NCOA3	PTK2B	RPL9	SPECC1L	ZDHHC5
ACTN2	CDKN2AIPNL	GNG5	NEB	PTPN12	RPP30	SRP14	ZFYVE27
ACTR2	CHAMP1	GOLM1	NGLY1	PUF60	RPS11	STX12	
ACTR3	CHST2	GPBP1	NLRP10	PUM2	RPS13	SULF2	
ADD1	CKAP2L	GSPT1	NLRP8	RAB29	RPS14	SUPT5H	
AGFG1	CLIC1	GSTO1	NOL8	RAPGEF3	RPS15	SZRD1	
AIF1L	CLSPN	H7C0S8	NOP56	RFC1	RPS15A	TAF7	
AIMP1	CNN1	HBA1	NPM1	RFX1	RPS17	TCERG1	
AKAP2	CSDE1	HBD	NPM3	RNMT	RPS18	TJP2	
ALCAM	CYR61	HDAC6	NR3C1	RPL10	RPS2	TK1	
ALDH3A2	DDX17	HINT2	NUDT1	RPL10A	RPS23	TLK1	
ANXA5	DDX3X	IARS2	NUMB	RPL11	RPS24	TM4SF1	
ANXA6	DFFA	IL17RA	NUMBL	RPL12	RPS26	TMEM154	
APEX1	DHRS2	IMMT	NUSAP1	RPL13A	RPS27	TNRC18	
APOE	DNMT1	KAT7	PA2G4	RPL15	RPS29	TNRC6B	
ARF4	DPY30	KHSRP	PALM	RPL18A	RPS3	TOMM34	
ARID1A	DYNLL2	KNSTRN	PALM3	RPL21	RPS3A	TPD52L2	
ARPC1B	ECHDC1	KPNA2	PDLIM2	RPL22	RPS4X	TSPAN31	
ATP6V1G1	EEF1D	LAP3	PDLIM5	RPL27	RPS8	TSPAN4	
BBX	EEF2	LEPRE1	PHGDH	RPL27A	RPS9	TUBA1B	
BOD1L1	EFTUD2	LETM2	POLR2G	RPL28	RPSA	TUBA1C	
C11orf98	EIF1AD	LIMD1	POP1	RPL30	RSRC2	TUBA3C	
C1orf122	EIF2S2	LRRFIP1	PPA1	RPL32	RTCB	TUBB2A	
C8orf88	EIF4A1	M6PR	PPA2	RPL34	S100A11	TUBB4B	
CALD1	EIF4EBP1	MBOAT7	PPP1R13L	RPL36	SACM1L	TUBB6	
CAP1	ELFN1	MCCC2	PPP1R14B	RPL36A	SEC11A	UBE2A	
CAPRIN1	ENAH	MCM3	PQBP1	RPL36AL	SEPHS2	USO1	
CAPZA1	ERBB2IP	MED15	PRDX6	RPL37A	SGOL2	WBP11	
CAPZB	ERH	MEF2D	PRMT5	RPL6	SLC4A1	WDR5	
CASKIN2	ERLIN2	MTCH1	PRRC2A	RPL7	SMARCAL1	XRN2	
CD3EAP	EXTL2	MYADM	PSMD4	RPL7A	SNW1	YWHAB	
CD81	FAM103A1	NAP1L1	PTGES3	RPL8	SPATA20	YWHAZ	

It was considered useful here to make one brief technical observation concerning nucleolar protein 8 (NOL8). There is a well-publicised and on-going problem with Microsoft Excel converting gene names into dates (Zeeberg et al. 2004; Ziemann et al. 2016). Whilst it is hoped that no errors survived proof-reading, this issue was a problem for the author in processing RNA data for chapter 5 whilst using an English language MacBook Pro. This current chapter was written on a Windows computer that is set up in the author's first national language, Gaeilge (Irish). It was noted that Microsoft Excel converted NOL8 into a date (Mollaig being the Irish

word for Christmas and *mí na Nollag* being the month of Christmas [i.e. December]). It is therefore likely to be the case that the problem with Excel is universal and that papers, whilst they may be written in English, may have different errors in them based on the language of choice of the authors. As was the case in chapter 5, it is hoped that any Microsoft Excel-induced errors were located and rectified prior to the use of the GO tool. This was, in fact, a main reason why it was decided to include every gene name exactly how it was entered into the GO tool in this chapter.

Table 6.41: Top ten overrepresented cellular component GO terms associated with proteins that were higher in abundance in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells without exclusions based on ANOVA p-value, by p-value

GO cellular component complete	Ref	Obs	Exp	FE	p-value
cytosolic ribosome (GO:0022626)	122	45	1.36	33.2	3.43E-50
ribosomal subunit (GO:0044391)	174	45	1.93	23.28	1.88E-43
adherens junction (GO:0005912)	684	68	7.6	8.95	3.48E-41
anchoring junction (GO:0070161)	702	68	7.8	8.72	1.76E-40
cytosolic part (GO:0044445)	243	47	2.7	17.41	7.67E-40
ribosome (GO:0005840)	252	47	2.8	16.79	3.92E-39
intracellular ribonucleoprotein complex (GO:0030529)	775	66	8.61	7.67	9.17E-36
ribonucleoprotein complex (GO:1990904)	776	66	8.62	7.66	9.91E-36
cell-substrate junction (GO:0030055)	400	50	4.44	11.25	1.13E-33
focal adhesion (GO:0005925)	393	49	4.37	11.22	7.08E-33

Whilst this was reassuring and implied that the analysis was sound, it was decided to undertake an additional analysis using the addition of extra proteins that were not considered significant to the 86 original hits (Table 6.33). It was hypothesised that the p-values for the overrepresented GO terms should become weaker as a result of the addition of non-significant proteins. It was decided to add the same number of additional proteins as was added following the widening of the selection universe (140). These were selected as the 140 least significant proteins by t-test p-value (either higher or lower in the arsenite-stressed si*UNR*-treated U2OS samples) when duplicated gene names were removed. Those 140 proteins (Table 6.42) were added to the 86 proteins previously accepted as being significantly down-regulated in arsenite-stressed si*UNR*-treated U2OS cells over arsenite-stressed control siRNA-treated U2OS cells (Table 6.33) and were fed into the GO tool to look for overrepresented cellular component GO-terms, the top ten of which by Bonferroni-adjusted p-value are presented in Table 6.43.

Table 6.42: List of non-significant genes/proteins added as a negative control to proteins that were significantly less abundant in arsenite-stressed si*UNR*-treated U2OS cells than in arsenite-stressed control siRNA-treated U2OS cells

ACBD5	CNN3	FKBP2	LAMC1	NUPL1	PUM2	SUOX
ACP2	COPS3	FTH1	LAMP1	ORC2	Q9UK23	TERF2
AHCY	COPS8	FXN	LDHA	P20290	Q9UPQ9	TGOLN2
AIMP1	CT45A9	GIGYF2	LGALS3	P28062	RAB10	TIPIN
AJUBA	CTTN	GNG2	LMAN2	P29692	RAB1B	TMEM106B
ALDOC	CYB5A	GOLIM4	LSM4	P40855	RAB7A	TMEM263
APOA1BP	DAD1	GPHN	LZIC	PCM1	RAD23B	TNFRSF12A
ARID2	DDA1	GPRIN1	MMTAG2	PCNP	RDX	TPM1
ARPC4-TTLL3	DIDO1	GPX1	MSN	PDLIM4	RLTPR	TPM3
BOLA3	DLST	HLA-A	MT-ATP8	PHB	RPS28	TSSC4
BPTF	DNASE2	HLA-B	MYL12B	PHF3	RRM2B	UBQLN4
C14orf119	DYNLL1	HMGB1	NACC1	PITHD1	SCAMP3	UBXN1
C9orf78	DYNLL2	HMGB3	NAPG	PKN2	SCPEP1	UQCRFS1
CCDC43	EIF2B4	HSPE1	NDUFV2	PPIF	SFPQ	VAPA
CD81	EMC6	HYPK	NFIC	PPP2CA	SGTB	VBP1
CFL1	ENG	ICT1	NME1	PPP4R2	SLC1A5	VKORC1L1
CFL2	ENO2	INIP	NME2	PRDX4	SMIM12	VTI1A
CHMP2B	ENSA	ITGA6	NUMA1	PRRC2A	SNAPIN	WARS
CNDP2	FAM192A	ITGB1	NUP50	PSMB7	SPTBN4	WNK1
CNN1	FBN2	KTN1	NUP62	PSMD9	STOML2	YWHAZ

Table 6.43: Top ten overrepresented cellular component GO terms associated with proteins that were higher in abundance in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells with 140 non-significant proteins added as a negative control

GO cellular component complete	REF	OBS	EXP	FE	p value
adherens junction (GO:0005912)	684	48	7.37	6.51	6.52E-22
anchoring junction (GO:0070161)	702	48	7.56	6.35	1.94E-21
extracellular exosome (GO:0070062)	2735	88	29.47	2.99	2.21E-19
extracellular vesicle (GO:1903561)	2749	88	29.62	2.97	3.12E-19
extracellular organelle (GO:0043230)	2750	88	29.63	2.97	3.20E-19
intracellular organelle (GO:0043229)	11822	192	127.4	1.51	4.08E-17
cytoplasm (GO:0005737)	10799	182	116.37	1.56	1.44E-16
macromolecular complex (GO:0032991)	4876	114	52.55	2.17	8.25E-16
extracellular region part (GO:0044421)	3825	99	41.22	2.4	9.81E-16
focal adhesion (GO:0005925)	393	32	4.24	7.56	1.61E-15

As expected, the p-values were lower than those that had been observed with the addition of proteins that were presumed to be significantly higher in abundance in arsenite-stressed control siRNA-treated U2OS cells than in unstressed si*UNR*-treated U2OS cells (Tables 6.41, 6.43). To compare the effect of widening the protein universe to accept additional significant proteins or adding additional non-significant proteins, it was decided to look at the top ten overrepresented cellular component GO-terms listed in Table 6.38. The fold enrichment values and p-values for each of the ten GO terms was then listed alongside the equivalent values where additional significant or non-significant proteins had been added (Table 6.44).

Table 6.44: Fold enrichment and p-value data for the top ten overrepresented cellular component GO terms associated with proteins that were higher in abundance in arsenite stressed control siRNA-treated U2OS cells than in arsenite stressed si*UNR*-treated U2OS cells, by p-value, with the addition of additional significant proteins (green), additional non-significant proteins (orange) or without any additional proteins (yellow)

	Original		+ significant		+ non-significant	
GO cellular component complete	**FE**	**p-value**	**FE**	**p-value**	**FE**	**p-value**
cytosolic ribosome (GO:0022626)	31.61	1.78E-16	33.2	3.43E-50	12.93	6.50E-11
adherens junction (GO:0005912)	9.16	4.25E-15	8.95	3.48E-41	6.51	6.52E-22
anchoring junction (GO:0070161)	8.93	7.93E-15	8.72	1.76E-40	6.35	1.94E-21
ribosomal subunit (GO:0044391)	22.17	4.42E-14	23.28	1.88E-43	9.60	1.61E-09
cytosolic part (GO:0044445)	16.86	3.62E-13	17.41	7.67E-40	8.44	6.01E-11
cytosol (GO:0005829)	3.32	4.21E-13	2.74	1.13E-21	2.47	3.20E-15
intracellular ribonucleoprotein complex (GO:0030529)	7.47	9.60E-12	7.67	9.17E-36	3.83	1.33E-07
ribonucleoprotein complex (GO:1990904)	7.46	9.87E-12	7.66	9.91E-36	3.83	1.37E-07
ribosome (GO:0005840)	15.31	1.29E-11	16.79	3.92E-39	6.63	6.06E-07
intracellular non-membrane-bounded organelle (GO:0043232)	2.84	4.64E-10	2.80	3.21E-28	2.07	1.53E-09

This showed that, for every GO term presented, the addition of the additional significant genes led to similar fold enrichments and p-values that were much stronger, by more than ten orders of magnitude for nine of the ten GO terms and by more than 20 orders of magnitude for eight of the ten (Table 6.44). These observations strongly support the suggestion that the significance detection criteria were stringent and that the observed hits are likely to be heavily populated with true positives. They also lend support to the idea that UNR may selectively modulate the expression of groups of proteins that are annotated to the GO terms that have been shown to

be overrepresented throughout this chapter. The addition of non-significant proteins, however, led to lower fold enrichments for all of the ten GO terms and weaker p-values for seven of the ten. This strongly supports the idea that the additional putative significant proteins were independently rich in true positives and that the putative non-significant proteins were indeed rich in false positives.

6.9.1 Consideration of GO term analysis with additional proteins

The observation that the p-values got better when more significant proteins were included was extremely reassuring as to the validity of the analysis. It should be noted that removal of proteins with high ANOVA p-values greatly improved the PCA analysis. It should also be noted, however, that ANOVA over two groups is similar to a t-test and, whilst paired t-tests ask a different question of groups of data to standard t-tests, it would nevertheless be expected that the paired t-tests alone would have led to a great improvement in the PCA analysis should that method have been offered by the Progenesis software.

6.10 Chapter summary

The chapter set out to explore the effect of knocking down UNR on protein expression under different conditions. The decision was made to use Progenesis or Scaffold to quantify mass spectrometry data and calculate an ANOVA p-value for each protein. Proteins with ANOVA p-values over 0.1 were excluded and paired t-tests were then carried out on the remaining proteins. Unfortunately, no individual proteins were particularly significant and none would have passed a multiple testing correction in which the total number of possible proteins (for example, 20000) was multiplied by the t-test p-value. The strongest p-value across all conditions was for prohibitin which was lower in si*UNR* treated HeLa cells than in control siRNA treated HeLa cells (p-value = 5.8×10^{-6}). Multiplying that p-value by 20,000 would just miss the p=0.1 cut off (adjusted p = 0.116). It was decided to add in proteins that either increased or decreased at least ten-fold on si*UNR* treatment, irrespective of their t-test p-value.

As the individual significant proteins were dubious as they would not pass a multiple testing correction, it was decided to use the Bonferroni multiple testing correction in GO term

overrepresentation analysis. It was assumed that, if the proteins were generally true hits, there may be some strong and reproducible overrepresented GO terms. If they were mainly random false positives, they would be unlikely to generate reproducible overrepresented GO terms with strong p-values. Many significantly overrepresented GO terms were found, including a number with very strong p-values. Some of these GO terms were reproducible between HeLa and U2OS and between arsenite stressed U2OS and unstressed U2OS. These observations led to the assumption that the methodology, as used, was sound.

One of the strongest and most reproducible GO terms was the molecular function GO term 'poly(A) RNA binding' which was a significant hit for the proteins that either increased or decreased on si*UNR* treatment of HeLa cells, as well as those that increased on si*UNR* treatment in unstressed U2OS cells or those that decreased on si*UNR* treatment in arsenite stressed U2OS cells. Possible the most reproducible of all the overrepresented GO terms was the cellular component GO term 'adherens junction'. This was significant in proteins that increased or decreased in both arsenite stressed and unstressed U2OS cells. It was also significant among proteins that increased following si*UNR* treatment in HeLa cells.

Having been satisfied that the methodology was adequate to locate differentially expressed proteins, it was then decided to see if it would be possible to find as many significant GO terms using extra proteins obtained by carrying out t-tests on all proteins, including those with ANOVA p-values greater than 0.1. This was tested using proteins that decreased in arsenite stressed si*UNR*-treated U2OS cells compared to arsenite stressed control siRNA-treated U2OS cells and looking at cellular component GO terms. This greatly improved the strength of the p-values (e.g. 'adherens junction' went from a Bonferroni-adjusted p-value of 4.25×10^{-15} to a Bonferroni-adjusted p-value of 3.48×10^{-41}). This observation provided even greater confidence in the results already obtained having excluded all proteins with ANOVA p-values over 0.1.

These results, as a minimum, support the claim that UNR modulates the expression of proteins involved in poly(A) binding and proteins that are localised to adherens junctions.

7 Summary and conclusions

The idea that there are two *UNR* genes in humans was suggested by Jeffers et al. (1990). No subsequent evidence was forthcoming to substantiate that claim. It was shown here that a region on Chromosome 10 had a similar degree of identity to Jeffers' probe as did the equivalent region on chromosome 1 that overlapped the human *UNR* gene. It is suggested here that Jeffers had mistakenly inferred that the region on chromosome 10 was a second *UNR* gene (section 1.1.5). A search of the human genome failed to locate a second *UNR* gene but did locate a previously discovered partial processed pseudogene of *UNR* on chromosome 5 (over 600 base pairs) and a stretch of over 300 base pairs on chromosome 7 that was almost 90% identical to part of the *UNR* gene that encodes a single exon (section 1.1.5).

It was noted that a disproportionately high number of arginine residues from the UNR protein are recorded as having at least one non-silent mutation in the COSMIC database (section 1.6.1).

Whilst it is not a novel finding, *per se*, the homemade ECL solution recipes stated in section 2.2.7 offer other labs the potential to save money compared to buying commercial ECL solutions. Whilst some optimisation was carried out within the lab, most credit is due to Dr Andrew Turnell for the suggestion and to Haan & Behrmann (2007) for the original research.

It was shown that UNR levels fall in cultured HeLa cells as they become more confluent (section 3.1.3) but this relationship is reversed in cultured U2OS cells (section 3.3.2).

It was shown that UNR and TP53 colocalise within stress granules in HeLa cells at one or two hours post-treatment with 1 mM sodium arsenite (section 3.1.6).

UNR levels were shown to be lower in 50%- or 70%-confluent U2OS cells that had been treated with 1 mM sodium arsenite for one hour compared to similar cells that were mock-treated with sterile PBS. This mirrored a similar reduction in TP53 levels (section 3.3.2).

Among the most reproducible novel UNR-interacting proteins discovered by RIP-mass spectrometry were the ubiquitin E3 ligase, HUWE1; the nucleolar protein, NARR; the multifunctional protein, SQSTM1; and the erythropoiesis-related protein, LDB1. Of these, HUWE1 was validated by Western blot and immunofluorescence microscopy and SQSTM1 was validated by Western blot. Each of these proteins would be interesting topics of future research. HUWE1 targets proteins involved in many processes, including cellular proliferation (e.g. MYC, which represses *UNR* transcription and has its own translation promoted by UNR) and promotes restart at stalled replication forks (Choe et al. 2016). Whilst the functional purpose of an interaction between UNR and HUWE1 is currently unknown, it is possible that UNR modulates HUWE1 function. Among other things, this would add an additional layer of complexity to the UNR-MYC relationship. As discussed below, UNR appears to regulate protein expression on multiple levels. It is further possible that interaction between UNR and a ubiquitin E3 ligase, such as HUWE1, could implicate UNR in protein turnover. Little work has been published on the NARR protein, which is the product of an alternative reading frame of an alternatively spliced *RAB34* transcript. It has been shown to be a nucleolar protein that is heavily phosphorylated during M phase (Zougman et al. 2011). UNR levels spike during M phase and this has been linked to the control of the cell cycle. It would be interesting to see if NARR and UNR play a joint role in the regulation of mitosis. SQSTM1 was shown to be involved in a number of interesting physiological pathways and pathologies, including autophagy (Katsuragi et al. 2015), tumour metastasis (Qiang et al. 2014) and Alzheimer's disease (Salminen et al. 2012). As UNR has also been reported to influence many of the same pathways as SQSTM1, it would be interesting to investigate whether or not the UNR-SQSTM1 interaction is important in respect to these processes. LDB1 has been shown to be involved in insulin expression (Hunter et al. 2013) and erythropoiesis (Li et al. 2010). UNR was shown to be differentially expressed in patients with uncontrolled diabetes (Xavier et al. 2014) and work by the von Lindern group in Holland has discovered links between UNR and Diamond Blackfan anaemia (Horos et al. 2012). As LDB1 influences transcription by LIM domain-containing transcription factors, an interaction between UNR and LDB1 could lead to UNR affecting pathways such as erythropoiesis and pathologies such as diabetes at the transcriptional level.

Groups of statistically significant putative UNR-interacting proteins from each cell type, either arsenite stressed or unstressed, were subjected to gene ontology overrepresentation analysis. The most overrepresented biological process GO term for all conditions was 'mRNA metabolic process' and the top two most overrepresented molecular function GO terms were 'RNA binding' and 'poly(A) RNA binding' (section 4.13). It was noted that 'ribonucleoprotein complex'

was a recurring highly overrepresented cellular component GO term (section 4.13). It is also suggested that UNR may be involved in selenium metabolism which was considered interesting in light of the findings of Xavier et al. (2014) (sections 4.13 and 1.7). Some evidence was found that UNR may bind to a number of proteins that are located at adherens junctions (e.g. section 4.10.5). This was interesting, given that some immunofluorescence microscopy evidence suggested that UNR can be found concentrated at cell-cell junctions under certain conditions (section 4.12.1). The adherens junction cellular component GO term was also significantly overrepresented among proteins that were upregulated in si*UNR*-treated HeLa cells or either unstressed or arsenite-stressed si*UNR*-treated U2OS cells (sections 6.4.1, 6.6.5 and 6.8.1). It was also overrepresented among proteins that were downregulated in either unstressed or arsenite-stressed si*UNR*-treated U2OS cells (sections 6.6.7 and 6.8.3). These findings suggested that UNR can both interact with proteins at adherens junctions and also modulate the expression levels of proteins at adherens junctions, making UNR a potential key player in the control of the adherens junction. This is suggested as a potential avenue of future research.

RNA was then extracted from a portion of the RIP samples and sequenced in an attempt to identify and quantify UNR-interacting transcripts. PCA was used to validate samples and remove outliers. DESeq2 was then used to find significant UNR-interacting transcripts. The significant hits were then entered into a GO term overrepresentation analysis tool to explore whether or not UNR interacts with specific groups of transcripts.

Multiple novel UNR-interacting transcripts are suggested in chapter 5, although there was little identity between the transcripts flagged up as binding to UNR in each of the three cell types. Notwithstanding that observation, some highly overrepresented GO terms were generated using DESeq2-suggested UNR-interacting transcripts merged from arsenite-stressed and unstressed SaOS-2 cells and from merged arsenite-stressed and unstressed cells from all three cell lines.

Among the top hits for SaOS-2 were ‘RNA processing’, ‘RNA binding’ and ‘poly(A) RNA binding’ (adjusted p-values ranging from 6.52×10^{-11} to 3.12×10^{-10}). The cellular component GO term ‘nuclear part’ had an adjusted p-value of 6.79×10^{-21} (section 5.6.2). Among all cell types, the most overrepresented biological process GO term was ‘nucleobase-containing compound metabolic process’ (adjusted p-value = 1.97×10^{-11}). Other top hits included ‘chromosome

organization' and 'chromatin modification' (adjusted p-values = $6.52x10^{-9}$ and $1.07x10^{-7}$, respectively). Among the most overrepresented molecular function GO terms were 'RNA binding' (adjusted p-value = $2.51x10^{-9}$) and 'poly (A) RNA binding' with an adjusted p-value of $1.08x10^{-6}$. The most overrepresented cellular component GO term was 'nuclear part' with an adjusted p-value of $4.51x10^{-23}$ (section 5.7.3). This evidence suggests that UNR interacts with transcripts that encode proteins that are involved in RNA- and nuclear-related processes. It is interesting to consider that UNR interacts with both proteins and transcripts that encode proteins involved in processes such as poly (A) RNA binding. This implies a potential fundamental role for UNR in the control of protein expression, potentially at the transcriptional, translational and post-translational levels.

A large number of proteins that are differentially regulated following siUnr-treatment in HeLa and U2OS cells (the latter with or without arsenite stress prior to harvesting) are presented in chapter 6. Akin to the analysis carried out with data from chapters 4 and 5, GO term overrepresentation analysis was carried out using these proteins.

Among the most overrepresented biological process GO terms generated using proteins upregulated in si*UNR*-treated HeLa cells were 'posttranscriptional regulation of gene expression' and terms related to viral processes. 26 proteins were observed that were annotated with 'poly (A) RNA binding', whereas only 7.07 were expected (adjusted p-value = $2.03x10^{-5}$). 'Extracellular exosome' and, as stated above, 'adherens junction', were among the most overrepresented cellular component GO terms (section 6.4.1). Among those proteins downregulated by si*UNR* treatment in HeLa cells, the most overrepresented biological process GO term was 'oxidative phosphorylation'. The top molecular function GO term was 'protein binding involved in cell-cell adhesion' and others included 'RNA binding' and 'poly (A) RNA binding'. Mitochondrion-associated GO terms were among the most overrepresented cellular component GO terms (section 6.4.3). These findings further corroborate the hypothesis that UNR has a fundamental role to play in the overall control of protein synthesis. They also suggest that UNR may play a role in cellular respiration.

The two most overrepresented biological process GO terms generated using proteins upregulated in unstressed si*UNR*-treated U2OS cells were 'mRNA processing' and 'RNA splicing'. 'Poly (A) RNA binding' was the top molecular function GO term with an adjusted p-value of

$5.11x10^{-12}$. 'Adherens junction' was among the most overrepresented cellular component GO terms, with 19 proteins observed with 3.26 expected (section 6.6.5).

Among the proteins downregulated by si*UNR* treatment in unstressed U2OS cells, there were no overrepresented biological process GO terms with adjusted p-values <0.01 and only two very general overrepresented molecular function GO terms. 'Anchoring junction' and 'extracellular exosome' were the most overrepresented cellular component GO terms (section 6.6.7).

There were neither any significantly overrepresented biological process nor molecular function GO terms obtained using proteins upregulated in si*UNR*-treated U2OS cells that were arsenite-stressed prior to being harvested. There were four significant cellular component GO terms, all of which had adjusted p-values >0.01, of which the most significant was 'adherens junction' (section 6.8.1).

A large number of proteins were significantly downregulated on si*UNR* treatment in U2OS cells that were stressed with arsenite prior to harvesting; 86 were entered into the GO tool (section 6.8.3). These yielded a large number of significantly overrepresented GO terms. The most significant biological process GO terms included terms related to translation and protein targeting. Another top ten hit by adjusted p-value was 'viral transcription'. 'RNA binding' (adjusted p-value = $2.79x10^{-17}$) and 'poly (A) RNA binding' (adjusted p-value = $5.69x10^{-14}$) were the most overrepresented molecular function GO terms by p-value. 'Cytosolic ribosome' (adjusted p-value = $1.78x10^{-16}$) and 'adherens junction' (adjusted p-value = $4.25x10^{-15}$) were the most overrepresented cellular component GO terms by p-value (section 6.8.3).

It was found that the method used to detect significant proteins among those discovered by mass spectrometry had been quite stringent as, when the criteria were relaxed, the p-values associated with significantly overrepresented GO terms became much stronger. The same was not true when putatively non-significant genes were added (section 6.9).

This work has produced evidence that supports a role for UNR as a potential master regulator of protein expression. The exact cellular function(s) of UNR remain(s) to be discovered but it is

hoped that this work will both widen the current knowledge of UNR and, potentially, help guide future researchers both in the UNR field and beyond.

7.1 Caveats

It is considered useful to conclude with a number of caveats. Firstly, mass spectrometry data can be biased in terms of differential detectability of processed tryptic digest peptides (Fricker 2015). The peptides that were detected here were fed into software that utilises databases, both of which are constantly changing and, thereby, changing the probabilities associated with protein detection. For example, the Scaffold report from section 4.3.1 mentions the use of version 4.5.3 that used Mascot version 2.5.0 and the human Uniprot database from 18th June 2015. As of 17th May 2017, Scaffold is on version 4.7.5, Mascot is on 2.6 and Uniprot had last updated its human proteome data on 25th March 2017. Ultimately, a parsimony-based approach was used to assign peptides to proteins and this is another potential source of error (Cottrell 2011). Likewise, RNA-Seq has many limitations from the sequencing depth to similar issues surrounding databases. Finally, gene ontology databases also change over time, both with respect to the number of terms and the group of terms annotated to each individual protein (Khatri & Draghici 2005). Possible errors are therefore magnified as false positive are likely to be present in the input data and some false negatives will have been rejected. This could potentially hide true associations or cause false relationships to be suggested. This is expected to have been particularly pronounced with the RNA-Seq data as it was not known whether a putative interaction with UNR would lead to a transcript being expressed at a greater rate or translationally repressed. Furthermore, the putative *UNR*-interacting transcripts were fed into a database that ultimately looked at protein attributes (biological process, molecular function and cellular component). It is hoped that a deeper exploration of the knockdown data provided in Chapter 6 could be used to strengthen future analysis of the RNA-Seq data from Chapter 5.

Acquaah-Mensah, G. K., Agu, N., Khan, T. & Gardner, A. (2015). A Regulatory Role for the Insulin- and BDNF-Linked RORA in the Hippocampus: Implications for Alzheimer's Disease. *J Alzheimer's Dis* **44**, 827–838. IOS Press.

Aebersold, R. & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* **422**, 198–207. Nature Publishing Group.

Anders, S., Pyl, P. T. & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169. Oxford University Press.

Anderson, E. C. & Ó Catnaigh, P. (2015). Regulation of the expression and activity of Unr in mammalian cells. *Biochem Soc Trans* **43**, 1241–1246. Portland Press Limited.

Anderson, E. C., Hunt, S. L. & Jackson, R. J. (2007). Internal initiation of translation from the human rhinovirus-2 internal ribosome entry site requires the binding of Unr to two distinct sites on the 5' untranslated region. *J Gen Virol* **88**, 3043–52.

Anumanthan, G., Halder, S. K., Friedman, D. B. & Datta, P. K. (2006). Oncogenic serine-threonine kinase receptor-associated protein modulates the function of Ewing sarcoma protein through a novel mechanism. *Cancer Res* **66**, 10824–32.

Araya, N., Hirota, K., Shimamoto, Y., Miyagishi, M., Yoshida, E., Ishida, J., Kaneko, S., Kaneko, M., Nakajima, T. & Fukamizu, A. (2003). Cooperative Interaction of EWS with CREB-binding Protein Selectively Activates Hepatocyte Nuclear Factor 4-mediated Transcription. *J Biol Chem* **278**, 5427–5432. American Society for Biochemistry and Molecular Biology.

Arnaoutov, A. & Dasso, M. (2003). The Ran GTPase Regulates Kinetochore Function. *Dev Cell* **5**, 99–111.

Azmi, A. S., Bao, B. & Sarkar, F. H. (2013). Exosomes in cancer development, metastasis, and drug resistance: a comprehensive review. *Cancer Metastasis Rev* **32**, 623–642. Springer US.

Bai, X., Kim, J., Yang, Z., Jurynec, M. J., Akie, T. E., Lee, J., LeBlanc, J., Sessa, A., Jiang, H. & other authors. **(2010).** TIF1gamma controls erythroid cell fate by regulating transcription elongation. *Cell* **142**, 133–43.

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209–11.

Beetz, C., Johnson, A., Schuh, A. L., Thakur, S., Varga, R.-E., Fothergill, T., Hertel, N., Bomba-Warczak, E., Thiele, H. & other authors. **(2013).** Inhibition of TFG function causes hereditary axon degeneration by impairing endoplasmic reticulum structure. *Proc Natl Acad Sci U S A* **110**, 5091–6.

Behjati, R., Kawai, K., Inadome, Y., Kano, J., Akaza, H. & Noguchi, M. (2011). APAF-1 is related to an undifferentiated state in the testicular germ cell tumor pathway. *Cancer Sci* **102**, 267–74.

Bell, D. W. (2014). Novel genetic targets in endometrial cancer. *Expert Opin Ther Targets* **18**, 725–30. NIH Public Access.

Berg, J. M. (1990). Zinc finger domains: hypotheses and current knowledge. *Annu Rev Biophys Biophys Chem* **19**, 405–21.

Berridge, M. J. (2014). Module 2: Cell Signalling Pathways. *Cell Signal Biol* **6**, csb0001002. Portland Press Journals portal.

Bessonnard, S., De Mot, L., Gonze, D., Barriol, M., Dennis, C., Goldbeter, A., Dupont, G. & Chazaud, C. (2014). Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* **141**.

Blechingberg, J., Luo, Y., Bolund, L., Damgaard, C. K. & Nielsen, A. L. (2012). Gene expression responses to FUS, EWS, and TAF15 reduction and stress granule sequestration analyses identifies FET-protein non-redundant functions. *PLoS One* **7**, e46251. Public Library of Science.

Bode, A. M. & Dong, Z. (2002). The paradox of arsenic: molecular mechanisms of cell transformation and chemotherapeutic effects. *Crit Rev Oncol Hematol* **42**, 5–24.

Boussadia, O., Amiot, F., Cases, S., Triqueneaux, G., Jacquemin-Sablon, H. & Dautry, F. (1997). Transcription of unr (upstream of N-ras) down-modulates N-ras expression in vivo. *FEBS Lett* **420**, 20–24.

Boussadia, O., Niepmann, M., Creancier, L., Prats, A.-C., Dautry, F. & Jacquemin-Sablon, H. (2003). Unr Is Required In Vivo for Efficient Initiation of Translation from the Internal Ribosome Entry Sites of both Rhinovirus and Poliovirus. *J Virol* **77**, 3353–3359.

Boussadia, O., Jacquemin-Sablon, H. & Dautry, F. (1993). Exon skipping in the expression of the gene immediately upstream of N-ras (unr/NRU). *Biochim Biophys Acta - Gene Struct Expr* **1172**, 64–72.

Bratton, S. B. & Salvesen, G. S. (2010). Regulation of the Apaf-1–caspase-9 apoptosome. *J Cell Sci* **123**.

Brown, E. C. & Jackson, R. J. (2004). All five cold-shock domains of unr (upstream of N-ras) are required for stimulation of human rhinovirus RNA translation. *J Gen Virol* **85**, 2279–2287.

Buchan, J. R. & Parker, R. (2009). Eukaryotic stress granules: the ins and outs of translation. *Mol Cell* **36**, 932–41.

Bullido, M. J., Martínez-García, A., Tenorio, R., Sastre, I., Muñoz, D. G., Frank, A. & Valdivieso, F. (2008). Double stranded RNA activated EIF2 α kinase (EIF2AK2; PKR) is associated with Alzheimer's disease. *Neurobiol Aging* **29**, 1160–1166.

Carthew, R. W. & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–55. NIH Public Access.

Caterina, M. J., Rosen, T. A., Tominaga, M., Brake, A. J. & Julius, D. (1999). A capsaicin-receptor homologue with a high threshold for noxious heat. *Nature* **398**, 436–41.

Cech, T. R. & Bass, B. L. (1986). Biological Catalysis by RNA. *Annu Rev Biochem* **55**, 599–629. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA .

Chang, K.-C., Chang, W.-C., Chang, Y., Hung, L.-Y., Lai, C.-H., Yeh, Y.-M., Chou, Y.-W. & Chen, C.-H. (2013). Ran GTPase-activating protein 1 is a therapeutic target in diffuse large B-cell lymphoma. *PLoS One* **8**, e79863. Public Library of Science.

Chang, T.-C., Yamashita, A., Chen, C.-Y. A., Yamashita, Y., Zhu, W., Durdan, S., Kahvejian, A., Sonenberg, N. & Shyu, A.-B. (2004). UNR, a new partner of poly(A)-binding protein, plays a key role in translationally coupled mRNA turnover mediated by the c-fos major coding-region determinant. *Genes Dev* **18**, 2010–23. Cold Spring Harbor Laboratory Press.

Chen, W. V, Delrow, J., Corrin, P. D., Frazier, J. P. & Soriano, P. (2004). Identification and validation of PDGF transcriptional targets by microarray-coupled gene-trap mutagenesis. *Nat Genet* **36**, 304–12.

Ching, T., Huang, S. & Garmire, L. X. (2014). Power analysis and sample size estimation for RNA-Seq differential expression. *RNA* **20**, 1684–1696. Cold Spring Harbor Laboratory Press.

Choe, K. N., Nicolae, C. M., Constantin, D., Imamura Kawasawa, Y., Delgado-Diaz, M. R., De, S., Freire, R., Smits, V. A. & Moldovan, G.-L. (2016). HUWE1 interacts with PCNA to alleviate replication stress. *EMBO Rep* e201541685. EMBO Press.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V & Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–40.

Christoph, F., Kempkensteffen, C., Weikert, S., Köllermann, J., Krause, H., Miller, K., Schostak, M. & Schrader, M. (2006). Methylation of tumour suppressor genes APAF-1 and DAPK-1 and in vitro effects of demethylating agents in bladder and kidney cancer. *Br J Cancer* **95**, 1701–7. Cancer Research UK.

Cléry, A., Blatter, M. & Allain, F. H.-T. (2008). RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* **18**, 290–298.

Coates, P. J., Jamieson, D. J., Smart, K., Prescott, A. R. & Hall, P. A. (1997). The prohibitin family of mitochondrial proteins regulate replicative lifespan. *Curr Biol*.

Coates, P. J., Nenutil, R., McGregor, A., Picksley, S. M., Crouch, D. H., Hall, P. A. & Wright, E. G. (2001). Mammalian Prohibitin Proteins Respond to Mitochondrial Stress and Decrease during Cellular Senescence. *Exp Cell Res* **265**, 262–273.

Cobbold, L. C., Wilson, L. A., Sawicka, K., King, H. A., Kondrashov, A. V, Spriggs, K. A., Bushell, M. & Willis, A. E. (2010). Upregulated c-myc expression in multiple myeloma by internal ribosome entry results from increased interactions with and expression of PTB-1 and YB-1. *Oncogene* **29**, 2884–2891. Nature Publishing Group.

Cohen, J. (1992). A power primer. *Psychol Bull* **112**, 155–9.

Coldwell, M. J., Mitchell, S. A., Stoneley, M., MacFarlane, M. & Willis, A. E. (2000). Initiation of Apaf-1 translation by internal ribosome entry. *Oncogene* **19**, 899–905.

Cornelis, S., Tinton, S. A., Schepens, B., Bruynooghe, Y. & Beyaert, R. (2005). UNR translation can be driven by an IRES element that is negatively regulated by polypyrimidine tract binding protein. *Nucleic Acids Res* **33**, 3095–108.

Da Costa, L., Moniz, H., Simansour, M., Tchernia, G., Mohandas, N. & Leblanc, T. (2010). Diamond-Blackfan anemia, ribosome and erythropoiesis. *Transfus Clin Biol* **17**, 112–119.

Costello, I., Nowotschin, S., Sun, X., Mould, A. W., Hadjantonakis, A.-K., Bikoff, E. K. & Robertson, E. J. (2015). Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. *Genes Dev* **29**, 2108–22.

Cottrell, J. S. (2011). Protein identification using MS/MS data. *J Proteomics* **74**, 1842–1851.

Crick, F. (1970). Central dogma of molecular biology. *Nature* **227**, 561–563.

Datta, P. K., Chytil, A., Gorska, A. E. & Moses, H. L. (1998). Identification of STRAP, a Novel WD Domain Protein in Transforming Growth Factor-beta Signaling. *J Biol Chem* **273**, 34671–34674. American Society for Biochemistry and Molecular Biology.

Datta, P. K. & Moses, H. L. (2000). STRAP and Smad7 synergize in the inhibition of transforming growth factor beta signaling. *Mol Cell Biol* **20**, 3157–67.

Davies, J., Denyer, T. & Hadfield, J. (2016). Bioanalyzer chips can be used interchangeably for many analyses of DNA or RNA. *Biotechniques* **60**, 197–199.

Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F. & Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235.

Dephoure, N., Zhou, C., Villén, J., Beausoleil, S. A., Bakalarski, C. E., Elledge, S. J. & Gygi, S. P. (2008). A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A* **105**, 10762–7.

Deuel, T. F. (2013). Anaplastic lymphoma kinase: 'Ligand Independent Activation' mediated by the PTN/RPTPβ/ζ signaling pathway. *Biochim Biophys Acta* **1834**, 2219–23.

Deveraux, Q. L., Takahashi, R., Salvesen, G. S. & Reed, J. C. (1997). X-linked IAP is a direct inhibitor of cell-death proteases. *Nature* **388**, 300–304. Nature Publishing Group.

Dhaka, A., Murray, A. N., Mathur, J., Earley, T. J., Petrus, M. J. & Patapoutian, A. (2007). TRPM8 Is Required for Cold Sensation in Mice. *Neuron* **54**, 371–378.

Diller, L., Kassel, J., Nelson, C. E., Gryka, M. A., Litwak, G., Gebhardt, M., Bressac, B., Ozturk, M., Baker, S. J. & Vogelstein, B. (1990). p53 functions as a cell cycle control protein in osteosarcomas. *Mol Cell Biol* **10**, 5772–81. American Society for Microbiology.

Dinur, M., Kilav, R., Sela-Brown, A., Jacquemin-Sablon, H. & Naveh-Many, T. (2006). In vitro evidence that upstream of N-ras participates in the regulation of parathyroid hormone messenger ribonucleic acid stability. *Mol Endocrinol Balt Md* **20**, 1652–1660.

Doniger, J. & DiPaolo, J. A. (1988). Coordinate N- ras mRNA up-regulation with mutational activation in tumorigenic guinea pig cells. *Nucleic Acids Res* **16**, 969–980.

Dormoy-Raclet, V., Markovits, J., Malato, Y., Huet, S., Lagarde, P., Montaudon, D., Jacquemin-Sablon, A. & Jacquemin-Sablon, H. (2007). Unr, a cytoplasmic RNA-binding protein with cold-shock domains, is involved in control of apoptosis in ES and HuH7 cells. *Oncogene* **26**, 2595–2605.

Dormoy-Raclet, V. (2005). *Caractérisation fonctionnelle de la protéine Unr (Upstream of N-ras) : rôle dans la réponse au stress et identification de gènes régulés par Unr*. Bordeaux 2.

Dormoy-Raclet, V., Markovits, J., Jacquemin-Sablon, A. & Jacquemin-Sablon, H. (2005). Regulation of Unr Expression by 5'- and 3'-Untranslated Regions of its mRNA through Modulation of Stability and IRES Mediated Translation. *RNA Biol* **2**, 112–120. Landes Bioscience.

Dorsam, R. T. & Kunapuli, S. P. (2004). Central role of the P2Y12 receptor in platelet activation. *J Clin Invest* **113**, 340–5. American Society for Clinical Investigation.

Dupont, S., Zacchigna, L., Cordenonsi, M., Soligo, S., Adorno, M., Rugge, M. & Piccolo, S. (2005). Germ-layer specification and control of cell growth by Ectodermin, a Smad4 ubiquitin ligase. *Cell* **121**, 87–99.

Dupont, S., Mamidi, A., Cordenonsi, M., Montagner, M., Zacchigna, L., Adorno, M., Martello, G., Stinchfield, M. J., Soligo, S. & other authors. **(2009).** FAM/USP9x, a deubiquitinating enzyme essential for TGFbeta signaling, controls Smad4 monoubiquitination. *Cell* **136**, 123–35.

Elatmani, H., Dormoy-Raclet, V., Dubus, P., Dautry, F., Chazaud, C. & Jacquemin-Sablon, H. (2011). The RNA-Binding Protein Unr Prevents Mouse Embryonic Stem Cells Differentiation Toward the Primitive Endoderm Lineage. *Stem Cells* **29**, 1504–16.

Elbashir, S., Harborth, J., Weber, K. & Tuschl, T. (2002). Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* **26**, 199–213.

Ender, C. & Meister, G. (2010). Argonaute proteins at a glance. *J Cell Sci* **123**.

Erkmann, J. A. & Kutay, U. (2004). Nuclear export of mRNA: from the site of transcription to the cytoplasm. *Exp Cell Res* **296**, 12–20.

Evans, J. R., Mitchell, S. A., Spriggs, K. A., Ostrowski, J., Bomsztyk, K., Ostarek, D. & Willis, A. E. (2003). Members of the poly (rC) binding protein family stimulate the activity of the c-myc internal ribosome entry segment in vitro and in vivo. *Oncogene* **22**, 8012–8020.

Feng, Y., Ariza, M. E., Goulet, A.-C., Shi, J. & Nelson, M. A. (2005). Death-signal-induced relocalization of cyclin-dependent kinase 11 to mitochondria. *Biochem J* **392**, 65–73. Portland Press Ltd.

Florenes, V. A., Maelandsmo, G. M., Forus, A., Andreassen, A., Myklebost, O. & Fodstad, O. (1994). MDM2 Gene Amplification and Transcript Levels in Human Sarcomas: Relationship to TP53 Gene Status. *JNCI J Natl Cancer Inst* **86**, 1297–1302. Oxford University Press.

Frankenberg, S., Gerbe, F., Bessonnard, S., Belville, C., Pouchin, P., Bardot, O., Chazaud, C., Arman, E., Haffner-Krausz, R. & other authors. **(2011).** Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Dev Cell* **21**, 1005–13. Elsevier.

Fricker, L. D. (2015). Limitations of Mass Spectrometry-Based Peptidomic Approaches. *J Am Soc Mass Spectrom* **26**, 1981–1991. Springer US.

Garcia, M. A., Gil, J., Ventoso, I., Guerra, S., Domingo, E., Rivas, C. & Esteban, M. (2006). Impact of Protein Kinase PKR in Cell Biology: from Antiviral to Antiproliferative Action. *Microbiol Mol Biol Rev* **70**, 1032–1060. American Society for Microbiology.

Ge, J., Apicella, M., Mills, J. A., Garçon, L., French, D. L., Weiss, M. J., Bessler, M. & Mason, P. J. (2015). Dysregulation of the Transforming Growth Factor β Pathway in Induced Pluripotent Stem Cells Generated from Patients with Diamond Blackfan Anemia. *PLoS One* **10**, e0134878. Public Library of Science.

Giaquinto, L., Curmi, P. M. G., Siddiqui, K. S., Poljak, A., DeLong, E., DasSarma, S. & Cavicchioli, R. (2007). Structure and function of cold shock proteins in archaea. *J Bacteriol* **189**, 5738–5748. American Society for Microbiology.

Gkika, D., Lemonnier, L., Shapovalov, G., Gordienko, D., Poux, C., Bernardini, M., Bokhobza, A., Bidaux, G., Degerny, C. & other authors. **(2015).** TRP channel-associated factors are a novel protein family that regulates TRPM8 trafficking and activity. *J Cell Biol* **208**, 89–107.

Gouzi, J. Y., Moog-Lutz, C., Vigny, M. & Brunet-de Carvalho, N. (2005). Role of the subcellular localization of ALK tyrosine kinase domain in neuronal differentiation of PC12 cells. *J Cell Sci* **118**, 5811–23.

Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* **36**, 59–74.

Graumann, P. L. & Marahiel, M. A. (1998). A superfamily of proteins that contain the cold-shock domain. *Trends Biochem Sci* **23**, 286–290.

Green, M. R. (1986). PRE-mRNA Splicing. *Annu Rev Genet* **20**, 671–708. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA .

Greer, S., Honeywell, R., Geletu, M., Arulanandam, R. & Raptis, L. (2010). Housekeeping genes; expression levels may change with density of cultured cells. *J Immunol Methods* **355**, 76–9.

Grimmler, M., Otter, S., Peter, C., Müller, F., Chari, A. & Fischer, U. (2005). Unrip, a factor implicated in cap-independent translation, associates with the cytosolic SMN complex and influences its intracellular localization. *Hum Mol Genet* **14**, 3099–111.

Grosset, C., Chen, C. Y., Xu, N., Sonenberg, N., Jacquemin-Sablon, H. & Shyu, A. B. (2000). A mechanism for translationally coupled mRNA turnover: interaction between the poly(A) tail and a c-fos RNA coding determinant via a protein complex. *Cell* **103**, 29–40.

Gu, L., Zhu, N., Zhang, H., Durden, D. L., Feng, Y. & Zhou, M. (2009). Regulation of XIAP Translation and Induction by MDM2 following Irradiation. *Cancer Cell* **15**, 363–375.

Gubitz, A. K., Feng, W. & Dreyfuss, G. (2004). The SMN complex. *Exp Cell Res* **296**, 51–56.

Haan, C. & Behrmann, I. (2007). A cost effective non-commercial ECL-solution for Western blot detections yielding strong signals and low background. *J Immunol Methods* **318**, 11–19.

Halder, S. K., Anumanthan, G., Maddula, R., Mann, J., Chytil, A., Gonzalez, A. L., Washington, M. K., Moses, H. L., Beauchamp, R. D. & Datta, P. K. (2006). Oncogenic function of a novel WD-domain protein, STRAP, in human carcinogenesis. *Cancer Res* **66**, 6156–66. American Association for Cancer Research.

Hall, T. M. T. (2005). Multiple modes of RNA recognition by zinc finger proteins. *Curr Opin Struct Biol* **15**, 367–373.

Hamilton, W. B., Brickman, J. M., Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J. K., Lonai, P., Artus, J., Piliszek, A. & other authors. **(2014).** Erk signaling suppresses embryonic stem cell self-renewal to specify endoderm. *Cell Rep* **9**, 2056–70. Elsevier.

Hanahan, D. & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674.

Hanahan, D. & Weinberg, R. A. (2000). The Hallmarks of Cancer. *Cell* **100**, 57–70.

Harbour, J. W. & Dean, D. C. (2000). The Rb/E2F pathway: expanding roles and emerging paradigms. *Genes Dev* **14**, 2393–2409. Cold Spring Harbor Laboratory Press.

Harima, Y., Togashi, A., Horikoshi, K., Imamura, M., Sougawa, M., Sawada, S., Tsunoda, T., Nakamura, Y. & Katagiri, T. (2004). Prediction of outcome of advanced cervical cancer to thermoradiotherapy according to expression profiles of 35 genes selected by cDNA microarray analysis. *Int J Radiat Oncol* **60**, 237–248.

He, L. & Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* **5**, 522–531. Nature Publishing Group.

He, W., Dorn, D. C., Erdjument-Bromage, H., Tempst, P., Moore, M. A. S. & Massagué, J. (2006). Hematopoiesis controlled by distinct TIF1gamma and Smad4 branches of the TGFbeta pathway. *Cell* **125**, 929–41.

Heidemann, J., Ogawa, H., Dwinell, M. B., Rafiee, P., Maaser, C., Gockel, H. R., Otterson, M. F., Ota, D. M., Lügering, N. & other authors. **(2003).** Angiogenic Effects of Interleukin 8 (CXCL8) in Human Intestinal Microvascular Endothelial Cells Are Mediated by CXCR2. *J Biol Chem* **278**, 8508–8515. American Society for Biochemistry and Molecular Biology.

Heo, I., Ha, M., Lim, J., Yoon, M.-J., Park, J.-E., Kwon, S. C., Chang, H. & Kim, V. N. (2012). Mono-Uridylation of Pre-MicroRNA as a Key Step in the Biogenesis of Group II let-7 MicroRNAs. *Cell* **151**, 521–532.

Holcik, M., Gordon, B. W. & Korneluk, R. G. (2003). The Internal Ribosome Entry Site-Mediated Translation of Antiapoptotic Protein XIAP Is Modulated by the Heterogeneous Nuclear Ribonucleoproteins C1 and C2. *Mol Cell Biol* **23**, 280–288.

Holt, C. E. & Bullock, S. L. (2009). Subcellular mRNA Localization in Animal Cells and Why It Matters. *Science (80-)* **326**.

Hoppe-Seyler, F. & Butz, K. (1993). Repression of endogenous p53 transactivation function in HeLa cervical carcinoma cells by human papillomavirus type 16 E6, human mdm-2, and mutant p53. *J Virol* **67**, 3111–3117.

Horos, R., Ijspeert, H., Pospisilova, D., Sendtner, R., Andrieu-Soler, C., Taskesen, E., Nieradka, A., Cmejla, R., Sendtner, M. & other authors. **(2012).** Ribosomal deficiencies in Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. *Blood* **119**, 262–72.

Howard, P. W., Ransom, D. G. & Maurer, R. A. (2010). Transcription intermediary factor 1gamma decreases protein expression of the transcriptional cofactor, LIM-domain-binding 1. *Biochem Biophys Res Commun* **396**, 674–8.

Hunt, S. L., Hsuan, J. J., Totty, N. & Jackson, R. J. (1999). unr, a cellular cytoplasmic RNA-binding protein with five cold-shock domains, is required for internal initiation of translation of human rhinovirus RNA. *Genes Dev* **13**, 437–448.

Hunter, C. S., Dixit, S., Cohen, T., Ediger, B., Wilcox, C., Ferreira, M., Westphal, H., Stein, R. & May, C. L. (2013). Islet α-, β-, and δ-cell development is controlled by the Ldb1 coregulator, acting primarily with the islet-1 transcription factor. *Diabetes* **62**, 875–86.

Hustinx, S. R., Cao, D., Maitra, A., Sato, N., Martin, S. T., Sudhir, D., Iacobuzio-Donahue, C., Cameron, J. L., Yeo, C. J. & other authors. **(2004).** Differentially expressed genes in pancreatic ductal adenocarcinomas identified through serial analysis of gene expression. *Cancer Biol Ther* **3**, 1254–61.

Ibrahim, F., Maragkakis, M., Alexiou, P., Maronski, M. A., Dichter, M. A. & Mourelatos, Z. (2013). Identification of in vivo, conserved, TAF15 RNA binding sites reveals the impact of TAF15 on the neuronal transcriptome. *Cell Rep* **3**, 301–8.

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Med* **2**, e124. Public Library of Science.

Ishigaki, S., Masuda, A., Fujioka, Y., Iguchi, Y., Katsuno, M., Shibata, A., Urano, F., Sobue, G. & Ohno, K. (2012). Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. *Sci Rep* **2**, 529. Nature Publishing Group.

Jackson, R. J. (2013). The current status of vertebrate cellular mRNA IRESs. *Cold Spring Harb Perspect Biol* **5**, a011569. Cold Spring Harbor Laboratory Press.

Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. **(2010).** The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113–27. Nature Publishing Group.

Jacquemin-Sablon, H., Triqueneaux, G., Deschamps, S., le Maire, M., Doniger, J. & Dautry, F. (1994a). Nucleic acid binding and intracellular localization of unr, a protein with five cold shock domains. *Nucleic Acids Res* **22**, 2643–2650.

Jacquemin-Sablon, H. & Dautry, F. (1992). Organization of the unrl N- ras locus: characterization of the promoter region of the human unr gene. *Nucleic Acids Res* **20**, 6355–6361.

Jacquemin-Sablon, H., Triqueneaux, G., Deschamps, S., le Maire, M., Doniger, J. & Dautry, F. (1994b). Nucleic acid binding and intracellular localization of unr, a protein with five cold shock domains. *Nucleic Acids Res* **22**, 2643–2650.

Jeffers, M., Paciucci, R. & Pellicer, A. (1990). Characterization of unr; a gene closely linked to N-ras. *Nucleic Acids Res* **18**, 4891–9.

Jia, L.-Q., Osada, M., Ishioka, C., Gamo, M., Ikawa, S., Suzuki, T., Shimodaira, H., Niitani, T., Kudo, T. & other authors. **(1997).** Screening thep53 status of human cell lines using a yeast functional assay. *Mol Carcinog* **19**, 243–253. Wiley Subscription Services, Inc., A Wiley Company.

Jiao, K., Zhou, Y. & Hogan, B. L. M. (2002). Identification of mZnf8, a mouse Krüppel-like transcriptional repressor, as a novel nuclear interaction partner of Smad1. *Mol Cell Biol* **22**, 7633–44.

Jonas, S. & Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* **16**, 421–433.

Jung, H., Seong, H.-A., Manoharan, R. & Ha, H. (2010). Serine-Threonine Kinase Receptor-associated Protein Inhibits Apoptosis Signal-regulating Kinase 1 Function through Direct Interaction. *J Biol Chem* **285**, 54–70. American Society for Biochemistry and Molecular Biology.

Kafasla, P., Morgner, N., Pöyry, T. A. A., Curry, S., Robinson, C. V & Jackson, R. J. (2009). Polypyrimidine tract binding protein stabilizes the encephalomyocarditis virus IRES structure via binding multiple sites in a unique orientation. *Mol Cell* **34**, 556–68.

Katsuragi, Y., Ichimura, Y. & Komatsu, M. (2015). p62/SQSTM1 functions as a signaling hub and an autophagy adaptor. *FEBS J* **282**, 4672–8.

Keene, J. D. (2007). RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet* **8**, 533–543. Nature Publishing Group.

Keene, J. D., Komisarow, J. M. & Friedersdorf, M. B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* **1**, 302–7. Nature Publishing Group.

Khatri, P. & Draghici, S. (2005). Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* **21**, 3587–3595. Oxford University Press.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S. L., Mortazavi, A., Williams, B., McCue, K. & other authors. **(2013).** TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36. BioMed Central.

Kim, J. H., Paek, K. Y., Choi, K., Kim, T.-D., Hahm, B., Kim, K.-T. & Jang, S. K. (2003). Heterogeneous Nuclear Ribonucleoprotein C Modulates Translation of c-myc mRNA in a Cell Cycle Phase-Dependent Manner. *Mol Cell Biol* **23**, 708–720. American Society for Microbiology.

Kim, M., Suh, J., Romano, D., Truong, M. H., Mullin, K., Hooli, B., Norton, D., Tesco, G., Elliott, K. & other authors. **(2009).** Potential late-onset Alzheimer's disease-associated mutations in the ADAM10 gene attenuate -secretase activity. *Hum Mol Genet* **18**, 3987–3996. Oxford University Press.

Kobayashi, H., Kawauchi, D., Hashimoto, Y., Ogata, T. & Murakami, F. (2013). The control of precerebellar neuron migration by RNA-binding protein Csde1. *Neuroscience* **253**, 292–303.

Kobayashi, H. & Tomari, Y. (2016). RISC assembly: Coordination between small RNAs and Argonaute proteins. *Biochim Biophys Acta - Gene Regul Mech* **1859**, 71–81.

Kojima, Y., Tam, O. H. & Tam, P. P. L. (2014). Timing of developmental events in the early mouse embryo. *Semin Cell Dev Biol* **34**, 65–75.

Kouhara, H., Hadari, Y. ., Spivak-Kroizman, T., Schilling, J., Bar-Sagi, D., Lax, I. & Schlessinger, J. (1997). A Lipid-Anchored Grb2-Binding Protein That Links FGF-Receptor Activation to the Ras/MAPK Signaling Pathway. *Cell* **89**, 693–702.

Kumar, R. & Thompson, J. R. (2011). The regulation of parathyroid hormone secretion and synthesis. *J Am Soc Nephrol* **22**, 216–24.

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680–685.

Lamant, L., Gascoyne, R. D., Duplantier, M. M., Armstrong, F., Raghab, A., Chhanabhai, M., Rajcan-Separovic, E., Raghab, J., Delsol, G. & Espinos, E. (2003). Non-muscle myosin heavy chain (MYH9): a new partner fused to ALK in anaplastic large cell lymphoma. *Genes Chromosomes Cancer* **37**, 427–32.

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359. Nature Publishing Group.

Lee, E. K., Kim, H. H., Kuwano, Y., Abdelmohsen, K., Srikantan, S., Subaran, S. S., Gleichmann, M., Mughal, M. R., Martindale, J. L. & other authors. **(2010).** hnRNP C promotes APP translation by competing with FMRP for APP mRNA recruitment to P bodies. *Nat Struct Mol Biol* **17**, 732–739. Nature Research.

Lee, S. B. & Esteban, M. (1994). The Interferon-induced Double-Stranded RNA-Activated Protein Kinase Induces Apoptosis. *Virology* **199**, 491–496. Academic Press.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O. & other authors. **(2003).** The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419.

Lenz, H. J., Leichman, C. G., Danenberg, K. D., Danenberg, P. V, Groshen, S., Cohen, H., Laine, L., Crookes, P., Silberman, H. & other authors. **(1996).** Thymidylate synthase mRNA level in adenocarcinoma of the stomach: a predictor for primary tumor response and overall survival. *J Clin Oncol* **14**, 176–82. American Society of Clinical Oncology.

Leppard, K. N. (2014). Adenoviruses: Molecular Biology. In *Ref Modul Biomed Sci*.

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. (1980). Are snRNPs involved in splicing? *Nature* **283**, 220–224. Nature Publishing Group.

Leshkowitz, D., Rozenblatt, O., Nakamura, T., Yano, T., Dautry, F., Croce, C. M. & Canaani, E. (1996). ALL-1 interacts with unr, a protein containing multiple cold shock domains. *Oncogene* **13**, 2027–31.

Li, L., Lee, J. Y., Gross, J., Song, S.-H., Dean, A. & Love, P. E. (2010). A requirement for Lim domain binding protein 1 in erythropoiesis. *J Exp Med* **207**, 2543–50.

Li, T., Inoue, A., Lahti, J. M. & Kidd, V. J. (2004). Failure To Proliferate and Mitotic Arrest of CDK11p110/p58-Null Mutant Mice at the Blastocyst Stage of Embryonic Cell Development. *Mol Cell Biol* **24**, 3188–3197.

Lin, J.-Y., Li, M.-L., Huang, P.-N., Chien, K.-Y., Horng, J.-T. & Shih, S.-R. (2008). Heterogeneous nuclear ribonuclear protein K interacts with the enterovirus 71 5’ untranslated region and participates in virus replication. *J Gen Virol* **89**, 2540–9.

Lin, J.-Y., Li, M.-L. & Shih, S.-R. (2009). Far upstream element binding protein 2 interacts with enterovirus 71 internal ribosomal entry site and negatively regulates viral translation. *Nucleic Acids Res* **37**, 47–59.

Lin, J.-Y., Li, M.-L. & Brewer, G. (2014). mRNA decay factor AUF1 binds the internal ribosomal entry site of enterovirus 71 and inhibits virus replication. *PLoS One* **9**, e103827 (L. Menéndez-Arias, Ed.). Public Library of Science.

Lin, L., Bass, A. J., Lockwood, W. W., Wang, Z., Silvers, A. L., Thomas, D. G., Chang, A. C., Lin, J., Orringer, M. B. & other authors. **(2012).** Activation of GATA binding protein 6 (GATA6) sustains oncogenic lineage-survival in esophageal adenocarcinoma. *Proc Natl Acad Sci U S A* **109**, 4251–6.

Lohrum, M. A. ., Ludwig, R. L., Kubbutat, M. H. ., Hanlon, M., Vousden, K. H., Ashcroft, M., Taya, Y., Vousden, K. H., Barak, Y. & other authors. **(2003).** Regulation of HDM2 activity by the ribosomal protein L11. *Cancer Cell* **3**, 577–587. Elsevier.

López-Fernández, L. A., López-Alañón, D. M. & del Mazo, J. (1995). Different developmental pattern of N-ras and unr gene expression in mouse gametogenic and somatic tissues. *Biochim Biophys Acta - Gene Struct Expr* **1263**, 10–16.

Loschi, M., Leishman, C. C., Berardone, N. & Boccaccio, G. L. (2009). Dynein and kinesin regulate stress-granule and P-body dynamics. *J Cell Sci* **122**, 3973–82.

Love, M. I., Huber, W., Anders, S., Lönnstedt, I., Speed, T., Robinson, M., Smyth, G., McCarthy, D., Chen, Y. & other authors. **(2014).** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550. BioMed Central.

Lu, Y. P., Ishiwata, T., Kawahara, K., Watanabe, M., Naito, Z., Moriyama, Y., Sugisaki, Y. & Asano, G. (2002). Expression of lumican in human colorectal cancer cells. *Pathol Int* **52**, 519–526. Blackwell Science Pty.

Luirink, J. & Sinning, I. (2004). SRP-mediated protein targeting: structure and function revisited. *Biochim Biophys Acta - Mol Cell Res* **1694**, 17–35.

Mao, D. Y. L., Watson, J. D., Yan, P. S., Barsyte-Lovejoy, D., Khosravi, F., Wong, W. W.-L., Farnham, P. J., Huang, T. H.-M. & Penn, L. Z. (2003). *Analysis of Myc Bound Loci Identified by CpG Island Arrays Shows that Max Is Essential for Myc-Dependent Repression*. *Curr Biol*.

Marchesini, M., Storti, P., Ogoti, Y., D'Anca, M., Nezi, L., Wei, Y., Yang, H., Ong, D., Neri, A. & other authors. **(2014).** ILF2 Is a Regulator of RNA Splicing and DNA Damage Response in 1q21-Amplified Multiple Myeloma. *Blood* **124**, 30. American Society of Hematology.

Mateo, E. C., Lorea, C. F., Duarte, A. A., Moreno, D., Neder, L., Junior, S. T., Scrileli, C. A. & Tone, L. G. (2011). A study of adrenocortical tumors in children and adolescents by a comparative genomic hybridization technique. *Cancer Genet* **204**, 298–308.

Matranga, C., Tomari, Y., Shin, C., Bartel, D. P. & Zamore, P. D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–20.

Matunis, M. J., Wu, J. & Blobel, G. (1998). SUMO-1 Modification and Its Role in Targeting the Ran GTPase-activating Protein, RanGAP1, to the Nuclear Pore Complex. *J Cell Biol* **140**, 499–509.

McGeachie, A. B., Koishi, K., Andrews, Z. B. & McLennan, I. S. (2005). Analysis of mRNAs that are enriched in the post-synaptic domain of the neuromuscular junction. *Mol Cell Neurosci* **30**, 173–185.

Mihailovich, M., Militti, C., Gabaldón, T. & Gebauer, F. (2010). Eukaryotic cold shock domain proteins: highly versatile regulators of gene expression. *BioEssays news Rev Mol Cell Dev Biol* **32**, 109–18.

Mihailovich, M., Wurth, L., Zambelli, F., Abaza, I., Militti, C., Mancuso, F. M., Roma, G., Pavesi, G. & Gebauer, F. (2012). Widespread generation of alternative UTRs contributes to sex-specific RNA binding by UNR. *RNA* **18**, 53–64.

Mitchell, S. A., Brown, E. C., Coldwell, M. J., Jackson, R. J. & Willis, A. E. (2001). Protein factor requirements of the Apaf-1 internal ribosome entry segment: roles of polypyrimidine tract binding protein and upstream of N-ras. *Mol Cell Biol* **21**, 3364–3374. American Society for Microbiology.

Mitchell, S. A., Spriggs, K. A., Coldwell, M. J., Jackson, R. J. & Willis, A. E. (2003). The Apaf-1 Internal Ribosome Entry Segment Attains the Correct Structural Conformation for Function via Interactions with PTB and unr. *Mol Cell* **11**, 757–771.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M. & Yamanaka, S. (2003). The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* **113**, 631–642.

Mizunuma, H., Miyazawa, J., Sanada, K. & Imai, K. (2003). The LIM-only protein, LMO4, and the LIM domain-binding protein, LDB1, expression in squamous cell carcinomas of the oral cavity. *Br J Cancer* **88**, 1543–8. Cancer Research UK.

Moallem, E., Silver, J. & Naveh-Many, T. (1995). Regulation of parathyroid hormone messenger RNA levels by protein kinase A and C in bovine parathyroid cells. *J Bone Miner Res* **10**, 447–52.

Moallem, E., Kilav, R., Silver, J. & Naveh-Many, T. (1998). RNA-Protein binding and post-transcriptional regulation of parathyroid hormone gene expression by calcium and phosphate. *J Biol Chem* **273**, 5253–5259.

Morrisey, E. E., Tang, Z., Sigrist, K., Lu, M. M., Jiang, F., Ip, H. S. & Parmacek, M. S. (1998). GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* **12**, 3579–3590. Cold Spring Harbor Laboratory Press.

Morsut, L., Yan, K.-P., Enzo, E., Aragona, M., Soligo, S. M., Wendling, O., Mark, M., Khetchoumian, K., Bressan, G. & other authors. **(2010).** Negative control of Smad activity by ectodermin/Tif1gamma patterns the mammalian embryo. *Development* **137**, 2571–8.

Motoori, M., Takemasa, I., Doki, Y., Saito, S., Miyata, H., Takiguchi, S., Fujiwara, Y., Yasuda, T., Yano, M. & other authors. **(2006).** Prediction of peritoneal metastasis in advanced gastric cancer by gene expression profiling of the primary site. *Eur J Cancer* **42**, 1897–1903.

Mukaida, N., Abe, S., Nakamura, H., Inoue, S., Takeda, H., Saito, H., Kato, S., Mukaida, N., Matsushima, K. & other authors. **(2003).** Pathophysiological roles of interleukin-8/CXCL8 in pulmonary diseases. *Am J Physiol Lung Cell Mol Physiol* **284**, L566-77. American Physiological Society.

Nagata, D., Suzuki, E., Nishimatsu, H., Yoshizumi, M., Mano, T., Walsh, K., Sata, M., Kakoki, M., Goto, A. & other authors. **(2000).** Cyclin A Downregulation and p21cip1 Upregulation Correlate With GATA-6-Induced Growth Arrest in Glomerular Mesangial Cells. *Circ Res* **87**, 699–704.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C. M. & Canaani, E. (2002). ALL-1 Is a Histone Methyltransferase that Assembles a Supercomplex of Proteins Involved in Transcriptional Regulation. *Mol Cell* **10**, 1119–1128.

Nakanishi, K. (2016). Anatomy of RISC: how do small RNAs and chaperones activate Argonaute proteins? *Wiley Interdiscip Rev RNA* **7**, 637–60. Wiley-Blackwell.

Nakayama, K., Nakayama, N., Wang, T.-L. & Shih, I.-M. (2007). NAC-1 Controls Cell Growth and Survival by Repressing Transcription of Gadd45GIP1, a Candidate Tumor Suppressor. *Cancer Res* **67**, 8058–8064.

Nakayama, K., Nakayama, N., Davidson, B., Sheu, J. J.-C., Jinawath, N., Santillan, A., Salani, R., Bristow, R. E., Morin, P. J. & other authors. **(2006).** A BTB/POZ protein, NAC-1, is related to tumor recurrence and is essential for tumor growth and survival. *Proc Natl Acad Sci U S A* **103**, 18739–44.

Nechama, M., Uchida, T., Mor Yosef-Levi, I., Silver, J. & Naveh-Many, T. (2009a). The peptidyl-prolyl isomerase Pin1 determines parathyroid hormone mRNA levels and stability in rat models of secondary hyperparathyroidism. *J Clin Invest* **119**, 3102–14.

Nechama, M., Peng, Y., Bell, O., Briata, P., Gherzi, R., Schoenberg, D. R. & Naveh-Many, T. (2009b). KSRP-PMR1-exosome association determines parathyroid hormone mRNA levels and stability in transfected cells. *BMC Cell Biol* **10**, 70.

Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646–58.

Ni, T., Mao, G., Xue, Q., Liu, Y., Chen, B., Cui, X., Lv, L., Jia, L., Wang, Y. & Ji, L. (2015). Upregulated expression of ILF2 in non-small cell lung cancer is associated with tumor cell proliferation and poor prognosis. *J Mol Histol* **46**, 325–35.

Niepmann, M. (1996). Porcine polypyrimidine tract-binding protein stimulates translation initiation at the internal ribosome entry site of foot-and-mouth-disease virus. *FEBS Lett* **388**, 39–42.

Niforou, K. N., Anagnostopoulos, A. K., Vougas, K., Kittas, C., Gorgoulis, V. G. & Tsangaris, G. T. (2006). The Proteome Profile of the Human Osteosarcome Saos2 Cell Line. *Cancer Genomics - Proteomics* **3**, 325–346. International Institute of Anticancer Research.

de Nijs, L., Léon, C., Nguyen, L., LoTurco, J. J., Delgado-Escueta, A. V, Grisar, T. & Lakaye, B. (2009). EFHC1 interacts with microtubules to regulate cell division and cortical development. *Nat Neurosci* **12**, 1266–1274. Nature Publishing Group.

O'Keefe, E. P. (2013). siRNAs and shRNAs: Tools for Protein Knockdown by Gene Silencing. *Mater Methods* **3**.

Olsen, J. V, Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S. & other authors. **(2010).** Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **3**, ra3.

Orivoli, S., Pavlidis, E., Cantalupo, G., Pezzella, M., Zara, F., Garavelli, L., Pisani, F. & Piccolo, B. (2016). Xp11.22 Microduplications Including HUWE1: Case Report and Literature Review. *Neuropediatrics* **47**, 51–6.

Osman, C., Merkwirth, C. & Langer, T. (2009). Prohibitins and the functional compartmentalization of mitochondrial membranes. *J Cell Sci* **122**, 3823–3830.

Othumpangat, S., Kashon, M. & Joseph, P. (2005). Sodium arsenite-induced inhibition of eukaryotic translation initiation factor 4E (eIF4E) results in cytotoxicity and cell death. *Mol Cell Biochem* **279**, 123–131.

Parisi, M. & Clayton, D. (1991). Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science (80-)* **252**.

Patel, G. P., Ma, S. & Bag, J. (2005). The autoregulatory translational control element of poly(A)-binding protein mRNA forms a heteromeric ribonucleoprotein complex. *Nucleic Acids Res* **33**, 7074–7089. Oxford University Press.

Payer, B. & Lee, J. T. (2008). X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**, 733–72.

Popat, S., Matakidou, A. & Houlston, R. S. (2004). Thymidylate synthase expression and prognosis in colorectal cancer: a systematic review and meta-analysis. *J Clin Oncol* **22**, 529–36. American Society of Clinical Oncology.

Potter, V. R. & Dubois, K. P. (1943). STUDIES ON THE MECHANISM OF HYDROGEN TRANSPORT IN ANIMAL TISSUES : VI. INHIBITOR STUDIES WITH SUCCINIC DEHYDROGENASE. *J Gen Physiol* **26**, 391–404. The Rockefeller University Press.

Qiang, L., Zhao, B., Ming, M., Wang, N., He, T.-C., Hwang, S., Thorburn, A. & He, Y.-Y. (2014). Regulation of cell proliferation and migration by p62 through stabilization of Twist1. *Proc Natl Acad Sci U S A* **111**, 9241–6.

Quarello, P., Garelli, E., Carando, A., Brusco, A., Calabrese, R., Dufour, C., Longoni, D., Misuraca, A., Vinti, L. & other authors. **(2010).** Diamond-Blackfan anemia: genotype-phenotype correlations in Italian patients with RPL5 and RPL11 mutations. *Haematologica* **95**.

Rahman, L., Voeller, D., Rahman, M., Lipkowitz, S., Allegra, C., Barrett, J. C., Kaye, F. J. & Zajac-Kaye, M. (2004). Thymidylate synthase as an oncogene: A novel role for an essential DNA synthesis enzyme. *Cancer Cell* **5**, 341–351.

Ransom, D. G., Bahary, N., Niss, K., Traver, D., Burns, C., Trede, N. S., Paffett-Lugassy, N., Saganic, W. J., Lim, C. A. & other authors. **(2004).** The zebrafish moonshine gene encodes transcriptional intermediary factor 1gamma, an essential regulator of hematopoiesis. *PLoS Biol* **2**, E237. Public Library of Science.

Ray, S. & Anderson, E. C. (2016). Stimulation of translation by human Unr requires cold shock domains 2 and 4, and correlates with poly(A) binding protein interaction. *Sci Rep* **6**, 22461. Nature Publishing Group.

Ray, S., Ó Catnaigh, P. & Anderson, E. C. (2015). Post-transcriptional regulation of gene expression by Unr. *Biochem Soc Trans* **43**, 323–327. Portland Press Limited.

Rein, K.-A., Borrebaek, B. & Bremer, J. (1979). Arsenite inhibits β-oxidation in isolated rat liver mitochondria. *Biochim Biophys Acta - Lipids Lipid Metab* **574**, 487–494.

Reiss, O. K. & Hellerman, L. (1958). Pyruvate utilization in heart sarcosomes; inhibition by an arsenoso compound and reactivation by lipoic acid. *J Biol Chem* **231**, 557–69.

Renner, S. (2014). Identification of ADAM10 5UTR binding proteins.

Romero-Calvo, I., Ocón, B., Martínez-Moya, P., Suárez, M. D., Zarzuelo, A., Martínez-Augustin, O. & de Medina, F. S. (2010). Reversible Ponceau staining as a loading control alternative to actin in Western blots. *Anal Biochem* **401**, 318–20.

Ruiz-Ramos, R., Lopez-Carrillo, L., Rios-Perez, A. D., De Vizcaya-Ruíz, A. & Cebrian, M. E. (2009). Sodium arsenite induces ROS generation, DNA oxidative damage, HO-1 and c-Myc proteins, NF-κB activation and cell proliferation in human breast cancer MCF-7 cells. *Mutat Res Toxicol Environ Mutagen* **674**, 109–115.

Sakamoto, S., Aoki, K., Higuchi, T., Todaka, H., Morisawa, K., Tamaki, N., Hatano, E., Fukushima, A., Taniguchi, T. & Agata, Y. (2009). The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. *Mol Cell Biol* **29**, 3754–69.

Salminen, A., Kaarniranta, K., Haapasalo, A., Hiltunen, M., Soininen, H. & Alafuzoff, I. (2012). Emerging role of p62/sequestosome-1 in the pathogenesis of Alzheimer's disease. *Prog Neurobiol* **96**, 87–95.

Santostefano, K. E., Hamazaki, T., Pardo, C. E., Kladde, M. P. & Terada, N. (2012). Fibroblast growth factor receptor 2 homodimerization rapidly reduces transcription of the pluripotency gene Nanog without dissociation of activating transcription factors. *J Biol Chem* **287**, 30507–17.

Savi, P., Labouret, C., Delesque, N., Guette, F., Lupker, J. & Herbert, J. M. (2001). P2y(12), a new platelet ADP receptor, target of clopidogrel. *Biochem Biophys Res Commun* **283**, 379–83.

Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J. & Howley, P. M. (1990). The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**, 1129–36.

Schepens, B., Tinton, S. A., Bruynooghe, Y., Parthoens, E., Haegman, M., Beyaert, R. & Cornelis, S. (2007). A role for hnRNP C1/C2 and Unr in internal initiation of translation during mitosis. *Eur Mol Biol Organ J* **26**, 158–169. Nature Publishing Group.

Schindelin, H., Marahiel, M. A. & Heinemann, U. (1993). Universal nucleic acid-binding domain revealed by crystal structure of the B. subtilis major cold-shock protein. *Nature* **364**, 164–8. Nature Publishing Group.

Schrode, N., Saiz, N., Di Talia, S. & Hadjantonakis, A.-K. (2014). GATA6 Levels Modulate Primitive Endoderm Cell Fate Choice and Timing in the Mouse Blastocyst. *Dev Cell* **29**, 454–467.

Schwartz, J. C., Cech, T. R. & Parker, R. R. (2015). Biochemical Properties and Biological Functions of FET Proteins. *Annu Rev Biochem* **84**, 355–79. Annual Reviews.

Sela-Brown, A., Silver, J., Brewer, G. & Naveh-Many, T. (2000). Identification of AUF1 as a Parathyroid Hormone mRNA 3'-Untranslated Region-binding Protein That Determines Parathyroid Hormone mRNA Stability. *J Biol Chem* **275**, 7424–7429.

Seong, H.-A., Jung, H., Ichijo, H. & Ha, H. (2010). Reciprocal Negative Regulation of PDK1 and ASK1 Signaling by Direct Interaction and Phosphorylation. *J Biol Chem* **285**, 2397–2414. American Society for Biochemistry and Molecular Biology.

Seong, H.-A., Jung, H., Choi, H.-S., Kim, K.-T. & Ha, H. (2005). Regulation of transforming growth factor-beta signaling and PDK1 kinase activity by physical interaction between PDK1 and serine-threonine kinase receptor-associated protein. *J Biol Chem* **280**, 42897–908.

Sester, M., Koebernick, K., Owen, D., Ao, M., Bromberg, Y., May, E., Stock, E., Andrews, L., Groh, V. & other authors. **(2010).** Conserved Amino Acids within the Adenovirus 2 E3/19K Protein Differentially Affect Downregulation of MHC Class I and MICA/B Proteins. *J Immunol* **184**, 255–267. American Association of Immunologists.

Setogawa, T., Shinozaki-Yabana, S., Masuda, T., Matsuura, K. & Akiyama, T. (2006). The tumor suppressor LKB1 induces p21 expression in collaboration with LMO4, GATA-6, and Ldb1. *Biochem Biophys Res Commun* **343**, 1186–90.

Shatsky, I. N., Dmitriev, S. E., Terenin, I. M. & Andreev, D. E. (2010). Cap- and IRES-independent scanning mechanism of translation initiation as an alternative to the concept of cellular IRESs. *Mol Cells* **30**, 285–93.

Shi, D. & Gu, W. (2012). Dual Roles of MDM2 in the Regulation of p53: Ubiquitination Dependent and Ubiquitination Independent Mechanisms of MDM2 Repression of p53 Activity. *Genes Cancer* **3**, 240–248. SAGE PublicationsSage CA: Los Angeles, CA.

Silver, J., Yalcindag, C., Sela-Brown, A., Kilav, R. & Naveh-Many, T. (1999). Regulation of the parathyroid hormone gene by vitamin D, calcium and phosphate. *Kidney Int* **56**, 2–7.

Snow, E., Sykora, P., Durham, T. & Klein, C. (2005). Arsenic, mode of action at biologically plausible low doses: What are the implications for low dose cancer risk? *Toxicol Appl Pharmacol* **207**, 557–564.

Soengas, M. S., Capodieci, P., Polsky, D., Mora, J., Esteller, M., Opitz-Araya, X., McCombie, R., Herman, J. G., Gerald, W. L. & other authors. **(2001).** Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. *Nature* **409**, 207–11. Macmillian Magazines Ltd.

Sonenberg, N. & Hinnebusch, A. G. (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**, 731–745.

Späth, G. F. & Weiss, M. C. (1998). Hepatocyte Nuclear Factor 4 Provokes Expression of Epithelial Marker Genes, Acting As a Morphogen in Dedifferentiated Hepatoma Cells. *J Cell Biol* **140**, 935–946. Rockefeller University Press.

Sum, E. Y. M., Peng, B., Yu, X., Chen, J., Byrne, J., Lindeman, G. J. & Visvader, J. E. (2002). The LIM domain protein LMO4 interacts with the cofactor CtIP and the tumor suppressor BRCA1 and inhibits BRCA1 activity. *J Biol Chem* **277**, 7849–56.

Sun, X., Fang, H., Li, X., Rossin, R., Welch, M. J. & Taylor, J.-S. (2005). MicroPET imaging of MCF-7 tumors in mice via unr mRNA-targeted peptide nucleic acids. *Bioconjug Chem* **16**, 294–305. American Chemical Society.

Takahama, K., Kino, K., Arai, S., Kurokawa, R. & Oyoshi, T. (2008). Identification of RNA binding specificity for the TET-family proteins. *Nucleic Acids Symp Ser (Oxf)* 213–4.

Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R. & other authors. **(2008).** Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534–538. Nature Publishing Group.

Tanabe, C., Hotoda, N., Sasagawa, N., Sehara-Fujisawa, A., Maruyama, K. & Ishiura, S. (2007). ADAM19 is tightly associated with constitutive Alzheimer's disease APP α-secretase in A172 cells. *Biochem Biophys Res Commun* **352**, 111–117.

Temin, H. M. & Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211–3.

Tinton, S. A., Schepens, B., Bruynooghe, Y., Beyaert, R. & Cornelis, S. (2005). Regulation of the cell-cycle-dependent internal ribosome entry site of the PITSLRE protein kinase: roles of Unr (upstream of N-ras) protein and phosphorylated translation initiation factor eIF-2alpha. *Biochem J* **385**, 155–63.

Tretyakova, I., Zolotukhin, A. S., Tan, W., Bear, J., Propst, F., Ruthel, G. & Felber, B. K. (2005). Nuclear export factor family protein participates in cytoplasmic mRNA trafficking. *J Biol Chem* **280**, 31981–90.

Triqueneaux, G., Velten, M., Franzon, P., Dautry, F. & Jacquemin-Sablon, H. (1999). RNA binding specificity of Unr, a protein with five cold shock domains. *Nucleic Acids Res* **27**, 1926–1934.

Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (80-)* **249**, 505–510. AAAS.

Turnell, A. (2008). Adenoviruses: Malignant Transformation and Oncology. In *Encycl Virol*, 3rd edn., pp. 9–16. Edited by & M. H. V. van R. B.W.J. Mahy. Oxford: Academic Press.

Versteeg, G. A., Rajsbaum, R., Sánchez-Aparicio, M. T., Maestre, A. M., Valdiviezo, J., Shi, M., Inn, K.-S., Fernandez-Sesma, A., Jung, J. & García-Sastre, A. (2013). The E3-Ligase TRIM Family of Proteins Regulates Signaling Pathways Triggered by Innate Immune Pattern-Recognition Receptors. *Immunity* **38**, 384–398.

Visvader, J. E., Venter, D., Hahm, K., Santamaria, M., Sum, E. Y., O'Reilly, L., White, D., Williams, R., Armes, J. & Lindeman, G. J. (2001). The LIM domain gene LMO4 inhibits differentiation of mammary epithelial cells in vitro and is overexpressed in breast cancer. *Proc Natl Acad Sci U S A* **98**, 14452–7.

Vukmirovic, M., Manojlovic, Z. & Stefanovic, B. (2013). Serine-threonine kinase receptor-associated protein (STRAP) regulates translation of type I collagen mRNAs. *Mol Cell Biol* **33**, 3893–906.

Wallace, D. M. & Cotter, T. G. (2009). Histone deacetylase activity in conjunction with E2F-1 and p53 regulates Apaf-1 expression in 661W cells and the retina. *J Neurosci Res* **87**, 887–905. Wiley Subscription Services, Inc., A Wiley Company.

Walton, R. J. (1977). A theoretical analysis of the steady-state relationship between plasma concentrations of calcium and parathyroid hormone. *Med Hypotheses* **3**, 1–3. Elsevier.

Wamaitha, S. E., del Valle, I., Cho, L. T. Y., Wei, Y., Fogarty, N. M. E., Blakeley, P., Sherwood, R. I., Ji, H. & Niakan, K. K. (2015). Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes Dev* **29**, 1239–55. Cold Spring Harbor Laboratory Press.

Wang, S., Tang, Y., Cui, H., Zhao, X., Luo, X., Pan, W., Huang, X. & Shen, N. (2011). Let-7/miR-98 regulate Fas and Fas-mediated apoptosis. *Genes Immun* **12**, 149–54. Macmillan Publishers Limited.

Waring, R. B. & Davies, R. W. (1984). Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing — a review. *Gene* **28**, 277–291.

White, J. P. & Lloyd, R. E. (2011). Poliovirus Unlinks TIA1 Aggregation and mRNA Stress Granule Formation. *J Virol* **85**, 12442–12454. American Society for Microbiology.

Wilson, R. C. & Doudna, J. A. (2013). Molecular Mechanisms of RNA Interference. *Annu Rev Biophys* **42**, 217–239. Annual Reviews .

Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359–362. Nature Publishing Group.

Wold, S., Esbensen, K. & Geladi, P. (1987). Principal component analysis. *Chemom Intell Lab Syst* **2**, 37–52. Elsevier.

Xavier, D. J., Takahashi, P., Manoel-Caetano, F. S., Foss-Freitas, M. C., Foss, M. C., Donadi, E. A., Passos, G. A. & Sakamoto-Hojo, E. T. (2014). One-week intervention period led to improvements in glycemic control and reduction in DNA damage levels in patients with type 2 diabetes mellitus. *Diabetes Res Clin Pract* **105**, 356–63.

Xia, K., Guo, H., Hu, Z., Xun, G., Zuo, L., Peng, Y., Wang, K., He, Y., Xiong, Z. & other authors. **(2013).** Common genetic variants on 1p13.2 associate with risk of autism. *Mol Psychiatry*. Macmillan Publishers Limited.

Xiao, Y.-C., Yang, Z.-B., Cheng, X.-S., Fang, X.-B., Shen, T., Xia, C.-F., Liu, P., Qian, H.-H., Sun, B. & other authors. **(2015).** CXCL8, overexpressed in colorectal cancer, enhances the resistance of colorectal cancer cells to anoikis. *Cancer Lett* **361**, 22–32. Elsevier.

Xu, B., Mao, Z., Ji, X., Yao, M., Chen, M., Zhang, X., Hang, B., Liu, Y., Tang, W. & other authors. **(2015).** miR-98 and its host gene Huwe1 target Caspase-3 in Silica nanoparticles-treated male germ cells. *Sci Rep* **5**, 12938. Nature Publishing Group.

Yang, J., Weinberg, R. A., Ahmed, S., Liu, C. C., Nawshad, A., Aigner, K., Dampier, B., Descovich, L., Mikula, M. & other authors. **(2008).** Epithelial-Mesenchymal Transition: At the Crossroads of Development and Tumor Metastasis. *Dev Cell* **14**, 818–829. Elsevier.

Yap, A. S., Brieher, W. M. & Gumbiner, B. M. (1997). MOLECULAR AND FUNCTIONAL ANALYSIS OF CADHERIN-BASED ADHERENS JUNCTIONS. *Annu Rev Cell Dev Biol* **13**, 119–146. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA .

Zermati, Y., Mouhamad, S., Stergiou, L., Besse, B., Galluzzi, L., Boehrer, S., Pauleau, A.-L., Rosselli, F., D'Amelio, M. & other authors. **(2007).** Nonapoptotic Role for Apaf-1 in the DNA Damage Checkpoint. *Mol Cell* **28**, 624–637.

Zhou, F., Liu, H., Zhang, X., Shen, Y., Zheng, D., Zhang, A., Lai, Y. & Li, H. (2014). Proline-rich protein 11 regulates epithelial-to-mesenchymal transition to promote breast cancer cell invasion. *Int J Clin Exp Pathol* **7**, 8692–9. e-Century Publishing Corporation.

Zhou, W., Fan, M.-Y., Wei, Y.-X., Huang, S., Chen, J.-Y. & Liu, P. (2016). The expression of MYH9 in osteosarcoma and its effect on the migration and invasion abilities of tumor cell. *Asian Pac J Trop Med* **In press**.

Zougman, A., Mann, M. & Wisniewski, J. R. (2011). Identification and characterization of a novel ubiquitous nucleolar protein 'NARR' encoded by a gene overlapping the rab34 oncogene. *Nucleic Acids Res* **39**, 7103–13.

Table S1: Putative Unr-interacting proteins in arsenite-stressed SaOS-2 cells, by t-test p-value (using Progenesis data – see section 4.11)

p-value	Protein
0.00025	PFKAL_HUMAN Isoform 2 of ATP-dependent 6-phosphofructokinase, liver type OS=Homo sapiens GN=PFKL
0.00052	SETMR_HUMAN Histone-lysine N-methyltransferase SETMAR OS=Homo sapiens GN=SETMAR PE=1 SV=2
0.00139	F5H2T0_HUMAN Elongator complex protein 1 OS=Homo sapiens GN=IKBKAP PE=1 SV=1
0.00170	A0A087WZL3_HUMAN Tyrosine-protein kinase receptor OS=Homo sapiens GN=ALK PE=1 SV=1
0.00241	WDR1_HUMAN WD repeat-containing protein 1 OS=Homo sapiens GN=WDR1 PE=1 SV=4
0.00272	PCBP1_HUMAN Poly(rC)-binding protein 1 OS=Homo sapiens GN=PCBP1 PE=1 SV=2
0.00312	CCD50_HUMAN Coiled-coil domain-containing protein 50 OS=Homo sapiens GN=CCDC50 PE=1 SV=1
0.00320	HSP76_HUMAN Heat shock 70 kDa protein 6 OS=Homo sapiens GN=HSPA6 PE=1 SV=2
0.00385	RINI_HUMAN Ribonuclease inhibitor OS=Homo sapiens GN=RNH1 PE=1 SV=2
0.00432	X6RK96_HUMAN TRMT1-like protein (Fragment) OS=Homo sapiens GN=TRMT1L PE=1 SV=1
0.00468	LDB1_HUMAN LIM domain-binding protein 1 OS=Homo sapiens GN=LDB1 PE=1 SV=2
0.00500	A0A075B6S5_HUMAN Protein IGKV1-27 (Fragment) OS=Homo sapiens GN=IGKV1-27 PE=1 SV=1
0.00531	ATRX_HUMAN Transcriptional regulator ATRX OS=Homo sapiens GN=ATRX PE=1 SV=5
0.00606	EF2_HUMAN Elongation factor 2 OS=Homo sapiens GN=EEF2 PE=1 SV=4
0.00700	J3KR97_HUMAN Tubulin-specific chaperone D OS=Homo sapiens GN=TBCD PE=1 SV=1
0.00731	A0A0A0MRH2_HUMAN Ryanodine receptor 1 OS=Homo sapiens GN=RYR1 PE=1 SV=1
0.00731	CO4A_HUMAN Complement C4-A OS=Homo sapiens GN=C4A PE=1 SV=2
0.00834	RS27A_HUMAN
0.00917	B4DXZ6_HUMAN Fragile X mental retardation syndrome-related protein 1 OS=Homo sapiens GN=FXR1 PE=1 SV=1
0.00986	A0A075B6K4_HUMAN HCG2043238 (Fragment) OS=Homo sapiens GN=IGLV3-10 PE=1 SV=1
0.01058	A0A0A0MTE2_HUMAN LIM domain only protein 7 OS=Homo sapiens GN=LMO7 PE=1 SV=1
0.01078	HS71A_HUMAN Heat shock 70 kDa protein 1A OS=Homo sapiens GN=HSPA1A PE=1 SV=1
0.01082	G5E9Q6_HUMAN Profilin OS=Homo sapiens GN=PFN2 PE=1 SV=1
0.01117	H0Y2W2_HUMAN ATPase family AAA domain-containing protein 3A (Fragment) OS=Homo sapiens GN=ATAD3A PE=1 SV=1
0.01130	SPG20_HUMAN Spartin OS=Homo sapiens GN=SPG20 PE=1 SV=1
0.01140	NARR_HUMAN Ras-related protein Rab-34, isoform NARR OS=Homo sapiens GN=RAB34 PE=1 SV=1
0.01197	ATD3A_HUMAN Isoform 2 of ATPase family AAA domain-containing protein 3A OS=Homo sapiens GN=ATAD3A
0.01248	P2Y12_HUMAN P2Y purinoceptor 12 OS=Homo sapiens GN=P2RY12 PE=1 SV=1
0.01249	TNR6C_HUMAN Trinucleotide repeat-containing gene 6C protein OS=Homo sapiens GN=TNRC6C PE=1 SV=3
0.01266	STRAP_HUMAN Serine-threonine kinase receptor-associated protein OS=Homo sapiens GN=STRAP PE=1 SV=1

0.01305	BAG2_HUMAN BAG family molecular chaperone regulator 2 OS=Homo sapiens GN=BAG2 PE=1 SV=1
0.01332	SSBP3_HUMAN Single-stranded DNA-binding protein 3 OS=Homo sapiens GN=SSBP3 PE=1 SV=1
0.01381	RS26L_HUMAN Putative 40S ribosomal protein S26-like 1 OS=Homo sapiens GN=RPS26P11 PE=5 SV=1
0.01408	H3BND4_HUMAN Pyridoxal-dependent decarboxylase domain-containing protein 1 OS=Homo sapiens GN=PDXDC1 PE=1 SV=1
0.01420	RBBP6_HUMAN E3 ubiquitin-protein ligase RBBP6 OS=Homo sapiens GN=RBBP6 PE=1 SV=1
0.01425	A0A087WY61_HUMAN Nuclear mitotic apparatus protein 1 OS=Homo sapiens GN=NUMA1 PE=1 SV=1
0.01436	E9PN89_HUMAN Heat shock cognate 71 kDa protein (Fragment) OS=Homo sapiens GN=HSPA8 PE=1 SV=1
0.01491	ATD3B_HUMAN ATPase family AAA domain-containing protein 3B OS=Homo sapiens GN=ATAD3B PE=1 SV=1
0.01494	CSDE1_HUMAN Isoform 3 of Cold shock domain-containing protein E1 OS=Homo sapiens GN=CSDE1
0.01495	E9PLT0_HUMAN Cold shock domain-containing protein E1 OS=Homo sapiens GN=CSDE1 PE=1 SV=1
0.01529	H3BRU6_HUMAN Poly(rC)-binding protein 2 (Fragment) OS=Homo sapiens GN=PCBP2 PE=1 SV=1
0.01530	H3BN98_HUMAN Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=2
0.01552	E9PKE3_HUMAN Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1
0.01558	E9PM36_HUMAN 40S ribosomal protein S2 OS=Homo sapiens GN=RPS2 PE=1 SV=1
0.01567	LMO4_HUMAN LIM domain transcription factor LMO4 OS=Homo sapiens GN=LMO4 PE=1 SV=1
0.01631	ATPG_HUMAN ATP synthase subunit gamma, mitochondrial OS=Homo sapiens GN=ATP5C1 PE=1 SV=1
0.01660	A0A0D9SFS2_HUMAN Thyroid receptor-interacting protein 6 OS=Homo sapiens GN=TRIP6 PE=4 SV=1
0.01670	M0QXS5_HUMAN Heterogeneous nuclear ribonucleoprotein L (Fragment) OS=Homo sapiens GN=HNRNPL PE=1 SV=1
0.01680	E9PQU5_HUMAN RNA-binding protein 25 (Fragment) OS=Homo sapiens GN=RBM25 PE=1 SV=1
0.01729	RS4X_HUMAN 40S ribosomal protein S4, X isoform OS=Homo sapiens GN=RPS4X PE=1 SV=2
0.01853	E9PLD4_HUMAN Cold shock domain-containing protein E1 (Fragment) OS=Homo sapiens GN=CSDE1 PE=1 SV=5
0.01854	CSDE1_HUMAN Cold shock domain-containing protein E1 OS=Homo sapiens GN=CSDE1 PE=1 SV=2
0.01856	V9GYM8_HUMAN Rho guanine nucleotide exchange factor 2 OS=Homo sapiens GN=ARHGEF2 PE=1 SV=1
0.01867	E7EMW7_HUMAN E3 ubiquitin-protein ligase UBR5 OS=Homo sapiens GN=UBR5 PE=1 SV=1
0.01892	E9PI65_HUMAN Heat shock cognate 71 kDa protein (Fragment) OS=Homo sapiens GN=HSPA8 PE=1 SV=1
0.01901	FND3B_HUMAN Fibronectin type III domain-containing protein 3B OS=Homo sapiens GN=FNDC3B PE=1 SV=2
0.01907	I3L404_HUMAN 40S ribosomal protein S2 (Fragment) OS=Homo sapiens GN=RPS2 PE=1 SV=1
0.01927	A0A0D9SF25_HUMAN F-box-like/WD repeat-containing protein TBL1XR1 (Fragment) OS=Homo sapiens GN=TBL1XR1 PE=4 SV=1
0.01929	F5H018_HUMAN GTP-binding nuclear protein Ran (Fragment) OS=Homo sapiens GN=RAN PE=1 SV=5
0.01958	G3V0J0_HUMAN Fragile X mental retardation 1, isoform CRA_e OS=Homo sapiens GN=FMR1 PE=1 SV=1
0.01991	TRIP6_HUMAN Thyroid receptor-interacting protein 6 OS=Homo sapiens GN=TRIP6 PE=1 SV=3

0.02029	H0YEN5_HUMAN 40S ribosomal protein S2 (Fragment) OS=Homo sapiens GN=RPS2 PE=1 SV=1
0.02045	J3KQE5_HUMAN GTP-binding nuclear protein Ran (Fragment) OS=Homo sapiens GN=RAN PE=1 SV=1
0.02048	TB10C_HUMAN Carabin OS=Homo sapiens GN=TBC1D10C PE=1 SV=1
0.02184	A0A087WVT6_HUMAN Single-stranded DNA-binding protein 3 OS=Homo sapiens GN=SSBP3 PE=1 SV=1
0.02200	Q5TFJ7_HUMAN Importin subunit alpha-7 (Fragment) OS=Homo sapiens GN=KPNA6 PE=1 SV=1
0.02317	HSP7C_HUMAN Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1
0.02352	J3KTE4_HUMAN Ribosomal protein L19 OS=Homo sapiens GN=RPL19 PE=1 SV=1
0.02444	E5RIM3_HUMAN Phospholipase A-2-activating protein OS=Homo sapiens GN=PLAA PE=1 SV=1
0.02474	E9PQD7_HUMAN 40S ribosomal protein S2 OS=Homo sapiens GN=RPS2 PE=1 SV=1
0.02498	C9JZR2_HUMAN Catenin delta-1 OS=Homo sapiens GN=CTNND1 PE=1 SV=2
0.02647	DYL2_HUMAN Dynein light chain 2, cytoplasmic OS=Homo sapiens GN=DYNLL2 PE=1 SV=1
0.02685	E7EMC7_HUMAN Sequestosome-1 OS=Homo sapiens GN=SQSTM1 PE=1 SV=1
0.02701	D6R9I7_HUMAN 40S ribosomal protein S23 OS=Homo sapiens GN=RPS23 PE=1 SV=1
0.02820	EFTU_HUMAN Elongation factor Tu, mitochondrial OS=Homo sapiens GN=TUFM PE=1 SV=2
0.03016	C9J2C0_HUMAN Tubulin alpha-8 chain (Fragment) OS=Homo sapiens GN=TUBA8 PE=1 SV=1
0.03092	TBB5_HUMAN Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2
0.03092	TBB4B_HUMAN Tubulin beta-4B chain OS=Homo sapiens GN=TUBB4B PE=1 SV=1
0.03238	XPOT_HUMAN Exportin-T OS=Homo sapiens GN=XPOT PE=1 SV=2
0.03248	SQSTM_HUMAN Sequestosome-1 OS=Homo sapiens GN=SQSTM1 PE=1 SV=1
0.03259	TBA1A_HUMAN Tubulin alpha-1A chain OS=Homo sapiens GN=TUBA1A PE=1 SV=1
0.03271	SYC2L_HUMAN Synaptonemal complex protein 2-like OS=Homo sapiens GN=SYCP2L PE=1 SV=2
0.03299	TBB2A_HUMAN Tubulin beta-2A chain OS=Homo sapiens GN=TUBB2A PE=1 SV=1
0.03317	RS27_HUMAN 40S ribosomal protein S27 OS=Homo sapiens GN=RPS27 PE=1 SV=3
0.03326	TTL12_HUMAN Tubulin--tyrosine ligase-like protein 12 OS=Homo sapiens GN=TTLL12 PE=1 SV=2
0.03535	F5H5D3_HUMAN Tubulin alpha-1C chain OS=Homo sapiens GN=TUBA1C PE=1 SV=1
0.03536	TBA4A_HUMAN Tubulin alpha-4A chain OS=Homo sapiens GN=TUBA4A PE=1 SV=1
0.03563	F8VQQ4_HUMAN Tubulin alpha-1A chain (Fragment) OS=Homo sapiens GN=TUBA1A PE=1 SV=1
0.03595	H3BLZ8_HUMAN Probable ATP-dependent RNA helicase DDX17 OS=Homo sapiens GN=DDX17 PE=1 SV=1
0.03653	TBA3C_HUMAN Tubulin alpha-3C/D chain OS=Homo sapiens GN=TUBA3C PE=1 SV=3
0.03696	Q5QNZ2_HUMAN ATP synthase F(0) complex subunit B1, mitochondrial OS=Homo sapiens GN=ATP5F1 PE=1 SV=1
0.03746	MCM7_HUMAN DNA replication licensing factor MCM7 OS=Homo sapiens GN=MCM7 PE=1 SV=4
0.03797	TBA1B_HUMAN Tubulin alpha-1B chain OS=Homo sapiens GN=TUBA1B PE=1 SV=1
0.03885	E7ET15_HUMAN U2 snRNP-associated SURP motif-containing protein OS=Homo sapiens GN=U2SURP PE=1 SV=1

0.03914	H0YBW4_HUMAN Phospholipase A-2-activating protein (Fragment) OS=Homo sapiens GN=PLAA PE=1 SV=1
0.03931	A0A0A0MS74_HUMAN Phosphoenolpyruvate carboxykinase [GTP], mitochondrial OS=Homo sapiens GN=PCK2 PE=1 SV=1
0.04050	F8WAE5_HUMAN Eukaryotic translation initiation factor 2A OS=Homo sapiens GN=EIF2A PE=1 SV=1
0.04057	PCKGM_HUMAN Phosphoenolpyruvate carboxykinase [GTP], mitochondrial OS=Homo sapiens GN=PCK2 PE=1 SV=3
0.04149	A0A075B736_HUMAN Tubulin beta-8 chain OS=Homo sapiens GN=TUBB8 PE=1 SV=1
0.04167	A0A087WX29_HUMAN TAR DNA-binding protein 43 (Fragment) OS=Homo sapiens GN=TARDBP PE=1 SV=1
0.04224	SYYC_HUMAN Tyrosine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=YARS PE=1 SV=4
0.04263	FAS_HUMAN Fatty acid synthase OS=Homo sapiens GN=FASN PE=1 SV=3
0.04314	RBM25_HUMAN RNA-binding protein 25 OS=Homo sapiens GN=RBM25 PE=1 SV=3
0.04322	G3V1A4_HUMAN Cofilin 1 (Non-muscle), isoform CRA_a OS=Homo sapiens GN=CFL1 PE=1 SV=1
0.04324	J3KTJ3_HUMAN 60S ribosomal protein L23 OS=Homo sapiens GN=RPL23 PE=1 SV=1
0.04339	H0Y368_HUMAN Dolichol-phosphate mannosyltransferase subunit 1 (Fragment) OS=Homo sapiens GN=DPM1 PE=1 SV=1
0.04472	H0YDD4_HUMAN Acetyltransferase component of pyruvate dehydrogenase complex (Fragment) OS=Homo sapiens GN=DLAT PE=1 SV=1
0.04491	M0QXL5_HUMAN rRNA 2'-O-methyltransferase fibrillarin (Fragment) OS=Homo sapiens GN=FBL PE=1 SV=1
0.04515	E3W990_HUMAN Sequestosome-1 (Fragment) OS=Homo sapiens GN=SQSTM1 PE=1 SV=1
0.04526	B4DEB1_HUMAN Histone H3 OS=Homo sapiens GN=H3F3A PE=1 SV=1
0.04557	SIR1_HUMAN NAD-dependent protein deacetylase sirtuin-1 OS=Homo sapiens GN=SIRT1 PE=1 SV=2
0.04732	J3KTM9_HUMAN Importin subunit beta-1 (Fragment) OS=Homo sapiens GN=KPNB1 PE=1 SV=1
0.04782	LR16A_HUMAN Leucine-rich repeat-containing protein 16A OS=Homo sapiens GN=LRRC16A PE=1 SV=1
0.04785	B0YJC4_HUMAN Vimentin OS=Homo sapiens GN=VIM PE=1 SV=1
0.04962	H0Y3Z3_HUMAN Protein disulfide-isomerase (Fragment) OS=Homo sapiens GN=P4HB PE=1 SV=1
0.05048	X6RLN4_HUMAN La-related protein 4 (Fragment) OS=Homo sapiens GN=LARP4 PE=1 SV=1
0.05188	COPA_HUMAN Coatomer subunit alpha OS=Homo sapiens GN=COPA PE=1 SV=2
0.05202	J3KT29_HUMAN 60S ribosomal protein L23 OS=Homo sapiens GN=RPL23 PE=1 SV=1
0.05246	IMB1_HUMAN Importin subunit beta-1 OS=Homo sapiens GN=KPNB1 PE=1 SV=2
0.05275	GRP78_HUMAN 78 kDa glucose-regulated protein OS=Homo sapiens GN=HSPA5 PE=1 SV=2
0.05414	H0YMR4_HUMAN Importin-4 (Fragment) OS=Homo sapiens GN=IPO4 PE=1 SV=1
0.05424	FBSP1_HUMAN F-box/SPRY domain-containing protein 1 OS=Homo sapiens GN=FBXO45 PE=1 SV=1
0.05424	NU107_HUMAN Nuclear pore complex protein Nup107 OS=Homo sapiens GN=NUP107 PE=1 SV=1
0.05479	SYAC_HUMAN Alanine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=AARS PE=1 SV=2
0.05534	G3V196_HUMAN Ras-related protein Rab-15 OS=Homo sapiens GN=RAB15 PE=1 SV=1
0.05640	DEST_HUMAN Destrin OS=Homo sapiens GN=DSTN PE=1 SV=3

0.05690	HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 OS=Homo sapiens GN=HUWE1 PE=1 SV=3
0.05730	P4HA2_HUMAN Prolyl 4-hydroxylase subunit alpha-2 OS=Homo sapiens GN=P4HA2 PE=1 SV=1
0.05801	XPF_HUMAN DNA repair endonuclease XPF OS=Homo sapiens GN=ERCC4 PE=1 SV=3
0.05913	H0YN14_HUMAN Importin-4 OS=Homo sapiens GN=IPO4 PE=1 SV=1
0.06011	IF2A_HUMAN Eukaryotic translation initiation factor 2 subunit 1 OS=Homo sapiens GN=EIF2S1 PE=1 SV=3
0.06481	K7ENJ4_HUMAN ATP synthase subunit alpha, mitochondrial (Fragment) OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
0.06544	I3L3H2_HUMAN Eukaryotic initiation factor 4A-III (Fragment) OS=Homo sapiens GN=EIF4A3 PE=1 SV=1
0.06572	E9PEJ4_HUMAN Acetyltransferase component of pyruvate dehydrogenase complex OS=Homo sapiens GN=DLAT PE=1 SV=1
0.06577	EDC4_HUMAN Enhancer of mRNA-decapping protein 4 OS=Homo sapiens GN=EDC4 PE=1 SV=1
0.06745	A0A0A0MSP3_HUMAN PDZ and LIM domain protein 5 OS=Homo sapiens GN=PDLIM5 PE=1 SV=1
0.06879	RS25_HUMAN 40S ribosomal protein S25 OS=Homo sapiens GN=RPS25 PE=1 SV=1
0.06886	IPO4_HUMAN Importin-4 OS=Homo sapiens GN=IPO4 PE=1 SV=2
0.06910	X1WI28_HUMAN 60S ribosomal protein L10 (Fragment) OS=Homo sapiens GN=RPL10 PE=1 SV=4
0.07024	E7END7_HUMAN Ras-related protein Rab-1A OS=Homo sapiens GN=RAB1A PE=1 SV=1
0.07066	E5RH77_HUMAN 40S ribosomal protein S14 OS=Homo sapiens GN=RPS14 PE=1 SV=1
0.07073	Q5T6W2_HUMAN Heterogeneous nuclear ribonucleoprotein K (Fragment) OS=Homo sapiens GN=HNRNPK PE=1 SV=1
0.07149	K7EK77_HUMAN ATP synthase subunit alpha, mitochondrial (Fragment) OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
0.07193	M0QX76_HUMAN 40S ribosomal protein S16 (Fragment) OS=Homo sapiens GN=RPS16 PE=1 SV=1
0.07256	RS14_HUMAN 40S ribosomal protein S14 OS=Homo sapiens GN=RPS14 PE=1 SV=3
0.07311	P4HA1_HUMAN Prolyl 4-hydroxylase subunit alpha-1 OS=Homo sapiens GN=P4HA1 PE=1 SV=2
0.07338	CY1_HUMAN Cytochrome c1, heme protein, mitochondrial OS=Homo sapiens GN=CYC1 PE=1 SV=3
0.07367	I3L397_HUMAN Eukaryotic translation initiation factor 5A (Fragment) OS=Homo sapiens GN=EIF5A PE=1 SV=5
0.07385	I3L1U9_HUMAN Actin, cytoplasmic 2 (Fragment) OS=Homo sapiens GN=ACTG1 PE=1 SV=1
0.07418	F8VPD4_HUMAN CAD protein OS=Homo sapiens GN=CAD PE=1 SV=1
0.07444	F5H423_HUMAN Uncharacterized protein OS=Homo sapiens PE=3 SV=1
0.07446	J3KMX5_HUMAN 40S ribosomal protein S13 OS=Homo sapiens GN=RPS13 PE=1 SV=1
0.07506	H7BY10_HUMAN 60S ribosomal protein L23a (Fragment) OS=Homo sapiens GN=RPL23A PE=1 SV=1
0.07558	F6RFD5_HUMAN Destrin OS=Homo sapiens GN=DSTN PE=1 SV=1
0.07624	J3JS69_HUMAN 40S ribosomal protein S18 OS=Homo sapiens GN=RPS18 PE=1 SV=1
0.07776	K7EM90_HUMAN Enolase (Fragment) OS=Homo sapiens GN=ENO1 PE=1 SV=1
0.07794	A2A3R7_HUMAN 40S ribosomal protein S6 OS=Homo sapiens GN=RPS6 PE=1 SV=1
0.07797	G5E9Y7_HUMAN LIM domain binding 2, isoform CRA_b OS=Homo sapiens GN=LDB2 PE=1 SV=1

0.07807	M0R0P1_HUMAN rRNA 2'-O-methyltransferase fibrillarin (Fragment) OS=Homo sapiens GN=FBL PE=1 SV=1
0.07865	A0A087WZ27_HUMAN Zinc finger protein 90 OS=Homo sapiens GN=ZNF90 PE=4 SV=2
0.07936	ATPA_HUMAN Isoform 3 of ATP synthase subunit alpha, mitochondrial OS=Homo sapiens GN=ATP5A1
0.07999	SERA_HUMAN D-3-phosphoglycerate dehydrogenase OS=Homo sapiens GN=PHGDH PE=1 SV=4
0.08021	HNRPK_HUMAN Heterogeneous nuclear ribonucleoprotein K OS=Homo sapiens GN=HNRNPK PE=1 SV=1
0.08302	I3L0K7_HUMAN Heat shock protein 75 kDa, mitochondrial OS=Homo sapiens GN=TRAP1 PE=1 SV=1
0.08349	A0A0D9SFL3_HUMAN RNA-binding protein EWS OS=Homo sapiens GN=EWSR1 PE=4 SV=1
0.08479	RS18_HUMAN 40S ribosomal protein S18 OS=Homo sapiens GN=RPS18 PE=1 SV=3
0.08523	Q5T4U5_HUMAN Acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain, isoform CRA_a OS=Homo sapiens GN=ACADM PE=1 SV=1
0.08630	ATPA_HUMAN ATP synthase subunit alpha, mitochondrial OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
0.08700	RS11_HUMAN 40S ribosomal protein S11 OS=Homo sapiens GN=RPS11 PE=1 SV=3
0.08985	Q5SZU1_HUMAN D-3-phosphoglycerate dehydrogenase OS=Homo sapiens GN=PHGDH PE=1 SV=1
0.09013	E9PLL6_HUMAN 60S ribosomal protein L27a OS=Homo sapiens GN=RPL27A PE=1 SV=1
0.09076	C9JJQ8_HUMAN Tubulin alpha-4A chain (Fragment) OS=Homo sapiens GN=TUBA4A PE=1 SV=1
0.09142	ZCCHV_HUMAN Zinc finger CCCH-type antiviral protein 1 OS=Homo sapiens GN=ZC3HAV1 PE=1 SV=3
0.09255	ACTB_HUMAN Actin, cytoplasmic 1 OS=Homo sapiens GN=ACTB PE=1 SV=1
0.09400	ARMC6_HUMAN Armadillo repeat-containing protein 6 OS=Homo sapiens GN=ARMC6 PE=1 SV=2
0.09452	KHDR2_HUMAN KH domain-containing, RNA-binding, signal transduction-associated protein 2 OS=Homo sapiens GN=KHDRBS2 PE=1 SV=1
0.09587	J3KNN5_HUMAN Probable ATP-dependent RNA helicase DDX41 (Fragment) OS=Homo sapiens GN=DDX41 PE=1 SV=1
0.09663	MYCB2_HUMAN E3 ubiquitin-protein ligase MYCBP2 OS=Homo sapiens GN=MYCBP2 PE=1 SV=3
0.09696	POTEE_HUMAN POTE ankyrin domain family member E OS=Homo sapiens GN=POTEE PE=1 SV=3
0.09733	SND1_HUMAN Staphylococcal nuclease domain-containing protein 1 OS=Homo sapiens GN=SND1 PE=1 SV=1

N.B. the t-test p-values are for a two-tailed paired t-test without multiple testing correction

Table S2: Putative Unr-interacting proteins in unstressed SaOS-2 cells, by Unr/IgG ratio (using Progenesis data – see section 4.11)

Unr/IgG ratio	Protein
∞	H3BLZ8_HUMAN Probable ATP-dependent RNA helicase DDX17 OS=Homo sapiens GN=DDX17 PE=1 SV=1
∞	TNR6C_HUMAN Trinucleotide repeat-containing gene 6C protein OS=Homo sapiens GN=TNRC6C PE=1 SV=3
∞	TRA2B_HUMAN Transformer-2 protein homolog beta OS=Homo sapiens GN=TRA2B PE=1 SV=1
∞	G5E9Y7_HUMAN LIM domain binding 2, isoform CRA_b OS=Homo sapiens GN=LDB2 PE=1 SV=1
∞	A0A087WY61_HUMAN Nuclear mitotic apparatus protein 1 OS=Homo sapiens GN=NUMA1 PE=1 SV=1
∞	C9JJQ8_HUMAN Tubulin alpha-4A chain (Fragment) OS=Homo sapiens GN=TUBA4A PE=1 SV=1
∞	H0Y3Z3_HUMAN Protein disulfide-isomerase (Fragment) OS=Homo sapiens GN=P4HB PE=1 SV=1
∞	H0Y9Y4_HUMAN 40S ribosomal protein S3a (Fragment) OS=Homo sapiens GN=RPS3A PE=1 SV=1
∞	BMP15_HUMAN Bone morphogenetic protein 15 OS=Homo sapiens GN=BMP15 PE=1 SV=2
∞	E9PKU4_HUMAN 60S ribosomal protein L8 (Fragment) OS=Homo sapiens GN=RPL8 PE=1 SV=1
∞	X6RAL5_HUMAN Histone deacetylase complex subunit SAP18 OS=Homo sapiens GN=SAP18 PE=1 SV=1
∞	AN32E_HUMAN Acidic leucine-rich nuclear phosphoprotein 32 family member E OS=Homo sapiens GN=ANP32E PE=1 SV=1
19575	AP2A2_HUMAN AP-2 complex subunit alpha-2 OS=Homo sapiens GN=AP2A2 PE=1 SV=2
3524	F8VVM2_HUMAN Phosphate carrier protein, mitochondrial OS=Homo sapiens GN=SLC25A3 PE=1 SV=1
2587	M0QXS5_HUMAN Heterogeneous nuclear ribonucleoprotein L (Fragment) OS=Homo sapiens GN=HNRNPL PE=1 SV=1
1662	D6R9I9_HUMAN ATP-binding cassette sub-family E member 1 OS=Homo sapiens GN=ABCE1 PE=1 SV=1
1299	H3BPE7_HUMAN RNA-binding protein FUS OS=Homo sapiens GN=FUS PE=1 SV=1
652	FBSP1_HUMAN F-box/SPRY domain-containing protein 1 OS=Homo sapiens GN=FBXO45 PE=1 SV=1
652	NU107_HUMAN Nuclear pore complex protein Nup107 OS=Homo sapiens GN=NUP107 PE=1 SV=1
511	V9GYZ6_HUMAN Bifunctional glutamate/proline--tRNA ligase (Fragment) OS=Homo sapiens GN=EPRS PE=1 SV=1
349	A0A0D9SF25_HUMAN F-box-like/WD repeat-containing protein TBL1XR1 (Fragment) OS=Homo sapiens GN=TBL1XR1 PE=4 SV=1
336	SSBP3_HUMAN Single-stranded DNA-binding protein 3 OS=Homo sapiens GN=SSBP3 PE=1 SV=1
336	A0A0D9SFL3_HUMAN RNA-binding protein EWS OS=Homo sapiens GN=EWSR1 PE=4 SV=1
298	E3W990_HUMAN Sequestosome-1 (Fragment) OS=Homo sapiens GN=SQSTM1 PE=1 SV=1
279	H0Y8N0_HUMAN Transmembrane protein 33 (Fragment) OS=Homo sapiens GN=TMEM33 PE=1 SV=1
204	HNRPK_HUMAN Heterogeneous nuclear ribonucleoprotein K OS=Homo sapiens GN=HNRNPK PE=1 SV=1
200	B4DLN1_HUMAN Uncharacterized protein OS=Homo sapiens PE=2 SV=1
199	B2R5W2_HUMAN Heterogeneous nuclear ribonucleoproteins C1/C2 OS=Homo sapiens GN=HNRNPC PE=1 SV=1
190	A0A087WVT6_HUMAN Single-stranded DNA-binding protein 3 OS=Homo sapiens GN=SSBP3 PE=1 SV=1
183	PSMD2_HUMAN 26S proteasome non-ATPase regulatory subunit 2 OS=Homo sapiens GN=PSMD2 PE=1 SV=3
173	E7ERJ7_HUMAN Polyadenylate-binding protein OS=Homo sapiens GN=PABPC1 PE=1 SV=1
168	B4DY09_HUMAN Interleukin enhancer-binding factor 2 OS=Homo sapiens GN=ILF2 PE=1 SV=1

165	LMO4_HUMAN LIM domain transcription factor LMO4 OS=Homo sapiens GN=LMO4 PE=1 SV=1
161	SQSTM_HUMAN Sequestosome-1 OS=Homo sapiens GN=SQSTM1 PE=1 SV=1
160	H7C086_HUMAN ATP synthase subunit O, mitochondrial (Fragment) OS=Homo sapiens GN=ATP5O PE=4 SV=1
149	E7EMC7_HUMAN Sequestosome-1 OS=Homo sapiens GN=SQSTM1 PE=1 SV=1
138	EDC4_HUMAN Enhancer of mRNA-decapping protein 4 OS=Homo sapiens GN=EDC4 PE=1 SV=1
137	Q5T6W2_HUMAN Heterogeneous nuclear ribonucleoprotein K (Fragment) OS=Homo sapiens GN=HNRNPK PE=1 SV=1
137	TTL12_HUMAN Tubulin--tyrosine ligase-like protein 12 OS=Homo sapiens GN=TTLL12 PE=1 SV=2
119	MCM7_HUMAN DNA replication licensing factor MCM7 OS=Homo sapiens GN=MCM7 PE=1 SV=4
112	J3KTE4_HUMAN Ribosomal protein L19 OS=Homo sapiens GN=RPL19 PE=1 SV=1
112	PO210_HUMAN Nuclear pore membrane glycoprotein 210 OS=Homo sapiens GN=NUP210 PE=1 SV=3
107	IMB1_HUMAN Importin subunit beta-1 OS=Homo sapiens GN=KPNB1 PE=1 SV=2
98	MPCP_HUMAN Phosphate carrier protein, mitochondrial OS=Homo sapiens GN=SLC25A3 PE=1 SV=2
90	LDB1_HUMAN LIM domain-binding protein 1 OS=Homo sapiens GN=LDB1 PE=1 SV=2
89	SRSF6_HUMAN Serine/arginine-rich splicing factor 6 OS=Homo sapiens GN=SRSF6 PE=1 SV=2
85	B4DR61_HUMAN Protein transport protein Sec61 subunit alpha isoform 1 OS=Homo sapiens GN=SEC61A1 PE=1 SV=1
83	COPA_HUMAN Coatomer subunit alpha OS=Homo sapiens GN=COPA PE=1 SV=2
80	CAND1_HUMAN Cullin-associated NEDD8-dissociated protein 1 OS=Homo sapiens GN=CAND1 PE=1 SV=2
77	I3L3H2_HUMAN Eukaryotic initiation factor 4A-III (Fragment) OS=Homo sapiens GN=EIF4A3 PE=1 SV=1
76	A0A087WTT1_HUMAN Polyadenylate-binding protein OS=Homo sapiens GN=PABPC1 PE=1 SV=1
73	NARR_HUMAN Ras-related protein Rab-34, isoform NARR OS=Homo sapiens GN=RAB34 PE=1 SV=1
68	J3KTM9_HUMAN Importin subunit beta-1 (Fragment) OS=Homo sapiens GN=KPNB1 PE=1 SV=1
64	LPPRC_HUMAN Leucine-rich PPR motif-containing protein, mitochondrial OS=Homo sapiens GN=LRPPRC PE=1 SV=3
61	ECHA_HUMAN Trifunctional enzyme subunit alpha, mitochondrial OS=Homo sapiens GN=HADHA PE=1 SV=2
57	RPN1_HUMAN Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 OS=Homo sapiens GN=RPN1 PE=1 SV=1
53	E7EMW7_HUMAN E3 ubiquitin-protein ligase UBR5 OS=Homo sapiens GN=UBR5 PE=1 SV=1
52	E9PN89_HUMAN Heat shock cognate 71 kDa protein (Fragment) OS=Homo sapiens GN=HSPA8 PE=1 SV=1
51	AT2A2_HUMAN Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 OS=Homo sapiens GN=ATP2A2 PE=1 SV=1
50	J3KQL8_HUMAN Apolipoprotein L2 OS=Homo sapiens GN=APOL2 PE=1 SV=2
48	E7EQU1_HUMAN High mobility group protein B3 (Fragment) OS=Homo sapiens GN=HMGB3 PE=1 SV=1
48	SYLC_HUMAN Leucine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=LARS PE=1 SV=2
44	A0A075B736_HUMAN Tubulin beta-8 chain OS=Homo sapiens GN=TUBB8 PE=1 SV=1
44	Q5QNZ2_HUMAN ATP synthase F(0) complex subunit B1, mitochondrial OS=Homo sapiens GN=ATP5F1 PE=1 SV=1
43	CCD50_HUMAN Coiled-coil domain-containing protein 50 OS=Homo sapiens GN=CCDC50 PE=1 SV=1
41	A2A3R7_HUMAN 40S ribosomal protein S6 OS=Homo sapiens GN=RPS6 PE=1 SV=1
40	E9PLD4_HUMAN Cold shock domain-containing protein E1 (Fragment) OS=Homo sapiens GN=CSDE1 PE=1 SV=5

40	H7BZ35_HUMAN Aspartate--tRNA ligase, cytoplasmic (Fragment) OS=Homo sapiens GN=DARS PE=1 SV=1
40	P4HA1_HUMAN Prolyl 4-hydroxylase subunit alpha-1 OS=Homo sapiens GN=P4HA1 PE=1 SV=2
38	M0R0P1_HUMAN rRNA 2'-O-methyltransferase fibrillarin (Fragment) OS=Homo sapiens GN=FBL PE=1 SV=1
38	F5H423_HUMAN Uncharacterized protein OS=Homo sapiens PE=3 SV=1
38	C9JZR2_HUMAN Catenin delta-1 OS=Homo sapiens GN=CTNND1 PE=1 SV=2
38	A0A0A0MSP3_HUMAN PDZ and LIM domain protein 5 OS=Homo sapiens GN=PDLIM5 PE=1 SV=1
37	A0A087WX29_HUMAN TAR DNA-binding protein 43 (Fragment) OS=Homo sapiens GN=TARDBP PE=1 SV=1
36	IF4A3_HUMAN Eukaryotic initiation factor 4A-III OS=Homo sapiens GN=EIF4A3 PE=1 SV=4
35	J3KMX5_HUMAN 40S ribosomal protein S13 OS=Homo sapiens GN=RPS13 PE=1 SV=1
35	H0YDD4_HUMAN Acetyltransferase component of pyruvate dehydrogenase complex (Fragment) OS=Homo sapiens GN=DLAT PE=1 SV=1
34	A6NLM8_HUMAN Translocon-associated protein subunit delta OS=Homo sapiens GN=SSR4 PE=1 SV=1
34	HSP76_HUMAN Heat shock 70 kDa protein 6 OS=Homo sapiens GN=HSPA6 PE=1 SV=2
33	PFKAL_HUMAN Isoform 2 of ATP-dependent 6-phosphofructokinase, liver type OS=Homo sapiens GN=PFKL
31	CNN3_HUMAN Calponin-3 OS=Homo sapiens GN=CNN3 PE=1 SV=1
31	CY1_HUMAN Cytochrome c1, heme protein, mitochondrial OS=Homo sapiens GN=CYC1 PE=1 SV=3
30	DYL2_HUMAN Dynein light chain 2, cytoplasmic OS=Homo sapiens GN=DYNLL2 PE=1 SV=1
30	F8VPD4_HUMAN CAD protein OS=Homo sapiens GN=CAD PE=1 SV=1
29	SRRM2_HUMAN Serine/arginine repetitive matrix protein 2 OS=Homo sapiens GN=SRRM2 PE=1 SV=2
29	ATD3A_HUMAN Isoform 2 of ATPase family AAA domain-containing protein 3A OS=Homo sapiens GN=ATAD3A
28	Q5T4U5_HUMAN Acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain, isoform CRA_a OS=Homo sapiens GN=ACADM PE=1 SV=1
28	E7EQV9_HUMAN Ribosomal protein L15 (Fragment) OS=Homo sapiens GN=RPL15 PE=1 SV=1
27	GNAL_HUMAN Guanine nucleotide-binding protein G(olf) subunit alpha OS=Homo sapiens GN=GNAL PE=1 SV=1
26	P2Y12_HUMAN P2Y purinoceptor 12 OS=Homo sapiens GN=P2RY12 PE=1 SV=1
26	CH60_HUMAN 60 kDa heat shock protein, mitochondrial OS=Homo sapiens GN=HSPD1 PE=1 SV=2
26	RAP1B_HUMAN Isoform 4 of Ras-related protein Rap-1b OS=Homo sapiens GN=RAP1B
26	A0A087X0X3_HUMAN Heterogeneous nuclear ribonucleoprotein M OS=Homo sapiens GN=HNRNPM PE=1 SV=1
25	P4HA2_HUMAN Prolyl 4-hydroxylase subunit alpha-2 OS=Homo sapiens GN=P4HA2 PE=1 SV=1
24	ADT2_HUMAN ADP/ATP translocase 2 OS=Homo sapiens GN=SLC25A5 PE=1 SV=7
24	M0QXL5_HUMAN rRNA 2'-O-methyltransferase fibrillarin (Fragment) OS=Homo sapiens GN=FBL PE=1 SV=1
23	DNJA1_HUMAN DnaJ homolog subfamily A member 1 OS=Homo sapiens GN=DNAJA1 PE=1 SV=2
23	X1WI28_HUMAN 60S ribosomal protein L10 (Fragment) OS=Homo sapiens GN=RPL10 PE=1 SV=4
22	E9PKE3_HUMAN Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1
21	SRRM1_HUMAN Serine/arginine repetitive matrix protein 1 OS=Homo sapiens GN=SRRM1 PE=1 SV=2
21	PCBP1_HUMAN Poly(rC)-binding protein 1 OS=Homo sapiens GN=PCBP1 PE=1 SV=2
21	CO4A_HUMAN Complement C4-A OS=Homo sapiens GN=C4A PE=1 SV=2
20	A0A0D9SEM4_HUMAN Serine/arginine-rich-splicing factor 4 (Fragment) OS=Homo sapiens GN=SRSF4 PE=4 SV=1

20	H0Y2W2_HUMAN ATPase family AAA domain-containing protein 3A (Fragment) OS=Homo sapiens GN=ATAD3A PE=1 SV=1
20	PCKGM_HUMAN Phosphoenolpyruvate carboxykinase [GTP], mitochondrial OS=Homo sapiens GN=PCK2 PE=1 SV=3
19	A0A0A0MS74_HUMAN Phosphoenolpyruvate carboxykinase [GTP], mitochondrial OS=Homo sapiens GN=PCK2 PE=1 SV=1
18	I3L0K7_HUMAN Heat shock protein 75 kDa, mitochondrial OS=Homo sapiens GN=TRAP1 PE=1 SV=1
18	TIF1B_HUMAN Transcription intermediary factor 1-beta OS=Homo sapiens GN=TRIM28 PE=1 SV=5
18	LR16A_HUMAN Leucine-rich repeat-containing protein 16A OS=Homo sapiens GN=LRRC16A PE=1 SV=1
18	HS90B_HUMAN Heat shock protein HSP 90-beta OS=Homo sapiens GN=HSP90AB1 PE=1 SV=4
18	J3KP15_HUMAN Serine/arginine-rich-splicing factor 2 (Fragment) OS=Homo sapiens GN=SRSF2 PE=1 SV=5
17	ATPA_HUMAN ATP synthase subunit alpha, mitochondrial OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
17	ATPB_HUMAN ATP synthase subunit beta, mitochondrial OS=Homo sapiens GN=ATP5B PE=1 SV=3
17	HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 OS=Homo sapiens GN=HUWE1 PE=1 SV=3
17	GRP75_HUMAN Stress-70 protein, mitochondrial OS=Homo sapiens GN=HSPA9 PE=1 SV=2
16	ATPA_HUMAN Isoform 3 of ATP synthase subunit alpha, mitochondrial OS=Homo sapiens GN=ATP5A1
16	KHDR1_HUMAN KH domain-containing, RNA-binding, signal transduction-associated protein 1 OS=Homo sapiens GN=KHDRBS1 PE=1 SV=1
16	SERA_HUMAN D-3-phosphoglycerate dehydrogenase OS=Homo sapiens GN=PHGDH PE=1 SV=4
16	RAGP1_HUMAN Ran GTPase-activating protein 1 OS=Homo sapiens GN=RANGAP1 PE=1 SV=1
16	J3KR97_HUMAN Tubulin-specific chaperone D OS=Homo sapiens GN=TBCD PE=1 SV=1
16	K7ENJ4_HUMAN ATP synthase subunit alpha, mitochondrial (Fragment) OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
16	XPOT_HUMAN Exportin-T OS=Homo sapiens GN=XPOT PE=1 SV=2
15	V9GYM8_HUMAN Rho guanine nucleotide exchange factor 2 OS=Homo sapiens GN=ARHGEF2 PE=1 SV=1
15	V9GYG0_HUMAN ADP/ATP translocase 1 OS=Homo sapiens GN=SLC25A4 PE=1 SV=1
15	K7EK77_HUMAN ATP synthase subunit alpha, mitochondrial (Fragment) OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
15	MYCB2_HUMAN E3 ubiquitin-protein ligase MYCBP2 OS=Homo sapiens GN=MYCBP2 PE=1 SV=3
15	NSD2_HUMAN Histone-lysine N-methyltransferase NSD2 OS=Homo sapiens GN=WHSC1 PE=1 SV=1
15	TBB5_HUMAN Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2
15	F8VQQ4_HUMAN Tubulin alpha-1A chain (Fragment) OS=Homo sapiens GN=TUBA1A PE=1 SV=1
14	EFTU_HUMAN Elongation factor Tu, mitochondrial OS=Homo sapiens GN=TUFM PE=1 SV=2
14	F8W026_HUMAN Endoplasmin (Fragment) OS=Homo sapiens GN=HSP90B1 PE=1 SV=5
14	ZCCHV_HUMAN Zinc finger CCCH-type antiviral protein 1 OS=Homo sapiens GN=ZC3HAV1 PE=1 SV=3
14	HSP7C_HUMAN Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1
14	ATD3B_HUMAN ATPase family AAA domain-containing protein 3B OS=Homo sapiens GN=ATAD3B PE=1 SV=1
14	E9PI65_HUMAN Heat shock cognate 71 kDa protein (Fragment) OS=Homo sapiens GN=HSPA8 PE=1 SV=1
14	HS90A_HUMAN Heat shock protein HSP 90-alpha OS=Homo sapiens GN=HSP90AA1 PE=1 SV=5
13	CSDE1_HUMAN Isoform 3 of Cold shock domain-containing protein E1 OS=Homo sapiens GN=CSDE1
13	GRP78_HUMAN 78 kDa glucose-regulated protein OS=Homo sapiens GN=HSPA5 PE=1 SV=2
13	POTEE_HUMAN POTE ankyrin domain family member E OS=Homo sapiens GN=POTEE PE=1 SV=3

13	G8JLB6_HUMAN Heterogeneous nuclear ribonucleoprotein H OS=Homo sapiens GN=HNRNPH1 PE=1 SV=1
13	ADT3_HUMAN ADP/ATP translocase 3 OS=Homo sapiens GN=SLC25A6 PE=1 SV=4
13	H2AZ_HUMAN Histone H2A.Z OS=Homo sapiens GN=H2AFZ PE=1 SV=2
13	TBB2A_HUMAN Tubulin beta-2A chain OS=Homo sapiens GN=TUBB2A PE=1 SV=1
13	H3BN98_HUMAN Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=2
13	RS27_HUMAN 40S ribosomal protein S27 OS=Homo sapiens GN=RPS27 PE=1 SV=3
13	H0Y368_HUMAN Dolichol-phosphate mannosyltransferase subunit 1 (Fragment) OS=Homo sapiens GN=DPM1 PE=1 SV=1
12	COPB_HUMAN Coatomer subunit beta OS=Homo sapiens GN=COPB1 PE=1 SV=3
12	TBB4B_HUMAN Tubulin beta-4B chain OS=Homo sapiens GN=TUBB4B PE=1 SV=1
12	DEST_HUMAN Destrin OS=Homo sapiens GN=DSTN PE=1 SV=3
12	J3KT29_HUMAN 60S ribosomal protein L23 OS=Homo sapiens GN=RPL23 PE=1 SV=1
12	H3BRG4_HUMAN Cytochrome b-c1 complex subunit 2, mitochondrial OS=Homo sapiens GN=UQCRC2 PE=1 SV=1
12	EF1A3_HUMAN Putative elongation factor 1-alpha-like 3 OS=Homo sapiens GN=EEF1A1P5 PE=5 SV=1
11	E9PLT0_HUMAN Cold shock domain-containing protein E1 OS=Homo sapiens GN=CSDE1 PE=1 SV=1
11	RL27_HUMAN 60S ribosomal protein L27 OS=Homo sapiens GN=RPL27 PE=1 SV=2
11	DYHC1_HUMAN Cytoplasmic dynein 1 heavy chain 1 OS=Homo sapiens GN=DYNC1H1 PE=1 SV=5
11	BAG2_HUMAN BAG family molecular chaperone regulator 2 OS=Homo sapiens GN=BAG2 PE=1 SV=1
11	E9PLL6_HUMAN 60S ribosomal protein L27a OS=Homo sapiens GN=RPL27A PE=1 SV=1
11	TBA1B_HUMAN Tubulin alpha-1B chain OS=Homo sapiens GN=TUBA1B PE=1 SV=1
11	A0A0A0MSG2_HUMAN Four and a half LIM domains protein 2 OS=Homo sapiens GN=FHL2 PE=1 SV=1
11	EF1A2_HUMAN Elongation factor 1-alpha 2 OS=Homo sapiens GN=EEF1A2 PE=1 SV=1
10	E7END7_HUMAN Ras-related protein Rab-1A OS=Homo sapiens GN=RAB1A PE=1 SV=1

N.B. Ratios are mean(Unr)/mean(IgG) and all infinities were ignored unless every Unr value was greater than zero. The yellow shading highlights a protein absent in every IgG sample but present in every Unr sample. Finite ratios were rounded to the nearest whole number.

Table S3: Putative Unr-interacting proteins in arsenite-stressed SaOS-2 cells, by Unr/IgG ratio (using Progenesis data – see section 4.11)

Unr/IgG ratio	Protein
∞	G5E9Q6_HUMAN Profilin OS=Homo sapiens GN=PFN2 PE=1 SV=1
∞	A0A087WY61_HUMAN Nuclear mitotic apparatus protein 1 OS=Homo sapiens GN=NUMA1 PE=1 SV=1
∞	TBR1_HUMAN T-box brain protein 1 OS=Homo sapiens GN=TBR1 PE=2 SV=1
∞	H3BLZ8_HUMAN Probable ATP-dependent RNA helicase DDX17 OS=Homo sapiens GN=DDX17 PE=1 SV=1
∞	AN32E_HUMAN Acidic leucine-rich nuclear phosphoprotein 32 family member E OS=Homo sapiens GN=ANP32E PE=1 SV=1
∞	TTL12_HUMAN Tubulin--tyrosine ligase-like protein 12 OS=Homo sapiens GN=TTLL12 PE=1 SV=2
∞	A0A0D9SF25_HUMAN F-box-like/WD repeat-containing protein TBL1XR1 (Fragment) OS=Homo sapiens GN=TBL1XR1 PE=4 SV=1
∞	FND3B_HUMAN Fibronectin type III domain-containing protein 3B OS=Homo sapiens GN=FNDC3B PE=1 SV=2
∞	FBSP1_HUMAN F-box/SPRY domain-containing protein 1 OS=Homo sapiens GN=FBXO45 PE=1 SV=1
∞	NU107_HUMAN Nuclear pore complex protein Nup107 OS=Homo sapiens GN=NUP107 PE=1 SV=1
∞	TNR6C_HUMAN Trinucleotide repeat-containing gene 6C protein OS=Homo sapiens GN=TNRC6C PE=1 SV=3
∞	SYLC_HUMAN Leucine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=LARS PE=1 SV=2
∞	MAGD1_HUMAN Melanoma-associated antigen D1 OS=Homo sapiens GN=MAGED1 PE=1 SV=3
∞	USMG5_HUMAN Up-regulated during skeletal muscle growth protein 5 OS=Homo sapiens GN=USMG5 PE=1 SV=1
∞	SUMO1_HUMAN
∞	F8WE65_HUMAN Peptidyl-prolyl cis-trans isomerase OS=Homo sapiens GN=PPIA PE=1 SV=1
∞	D6R9I9_HUMAN ATP-binding cassette sub-family E member 1 OS=Homo sapiens GN=ABCE1 PE=1 SV=1
∞	BMP15_HUMAN Bone morphogenetic protein 15 OS=Homo sapiens GN=BMP15 PE=1 SV=2
∞	U3KQF2_HUMAN ADP-ribosylation factor 4 OS=Homo sapiens GN=ARF4 PE=1 SV=1
∞	H0Y8R5_HUMAN Guanine nucleotide-binding protein subunit beta-2-like 1 (Fragment) OS=Homo sapiens GN=GNB2L1 PE=1 SV=1
∞	H0Y9Y4_HUMAN 40S ribosomal protein S3a (Fragment) OS=Homo sapiens GN=RPS3A PE=1 SV=1
∞	J3KP15_HUMAN Serine/arginine-rich-splicing factor 2 (Fragment) OS=Homo sapiens GN=SRSF2 PE=1 SV=5
∞	B5MCT8_HUMAN 40S ribosomal protein S9 OS=Homo sapiens GN=RPS9 PE=1 SV=1
∞	G5E9Y7_HUMAN LIM domain binding 2, isoform CRA_b OS=Homo sapiens GN=LDB2 PE=1 SV=1
16177	X6RLN4_HUMAN La-related protein 4 (Fragment) OS=Homo sapiens GN=LARP4 PE=1 SV=1
15686	SSBP3_HUMAN Single-stranded DNA-binding protein 3 OS=Homo sapiens GN=SSBP3 PE=1 SV=1
2605	Q5T6W2_HUMAN Heterogeneous nuclear ribonucleoprotein K (Fragment) OS=Homo sapiens GN=HNRNPK PE=1 SV=1
1805	A0A0A0MSP3_HUMAN PDZ and LIM domain protein 5 OS=Homo sapiens GN=PDLIM5 PE=1 SV=1
1779	SYYC_HUMAN Tyrosine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=YARS PE=1 SV=4
1401	H3BM89_HUMAN 60S ribosomal protein L4 OS=Homo sapiens GN=RPL4 PE=1 SV=1
1354	B4DY09_HUMAN Interleukin enhancer-binding factor 2 OS=Homo sapiens GN=ILF2 PE=1 SV=1
1317	F8VVM2_HUMAN Phosphate carrier protein, mitochondrial OS=Homo sapiens GN=SLC25A3 PE=1 SV=1
637	E7EQU1_HUMAN High mobility group protein B3 (Fragment) OS=Homo sapiens GN=HMGB3 PE=1 SV=1
553	SQSTM_HUMAN Sequestosome-1 OS=Homo sapiens GN=SQSTM1 PE=1 SV=1

550	DEST_HUMAN Destrin OS=Homo sapiens GN=DSTN PE=1 SV=3
512	E9PLD4_HUMAN Cold shock domain-containing protein E1 (Fragment) OS=Homo sapiens GN=CSDE1 PE=1 SV=5
465	EDC4_HUMAN Enhancer of mRNA-decapping protein 4 OS=Homo sapiens GN=EDC4 PE=1 SV=1
456	M0R0P1_HUMAN rRNA 2'-O-methyltransferase fibrillarin (Fragment) OS=Homo sapiens GN=FBL PE=1 SV=1
447	E3W990_HUMAN Sequestosome-1 (Fragment) OS=Homo sapiens GN=SQSTM1 PE=1 SV=1
366	E7ERJ7_HUMAN Polyadenylate-binding protein OS=Homo sapiens GN=PABPC1 PE=1 SV=1
339	F8VPD4_HUMAN CAD protein OS=Homo sapiens GN=CAD PE=1 SV=1
286	A0A087WX29_HUMAN TAR DNA-binding protein 43 (Fragment) OS=Homo sapiens GN=TARDBP PE=1 SV=1
278	E7EMC7_HUMAN Sequestosome-1 OS=Homo sapiens GN=SQSTM1 PE=1 SV=1
278	F8WAE5_HUMAN Eukaryotic translation initiation factor 2A OS=Homo sapiens GN=EIF2A PE=1 SV=1
273	C9JJQ8_HUMAN Tubulin alpha-4A chain (Fragment) OS=Homo sapiens GN=TUBA4A PE=1 SV=1
261	A0A087WTT1_HUMAN Polyadenylate-binding protein OS=Homo sapiens GN=PABPC1 PE=1 SV=1
249	C9J7E5_HUMAN Transportin-3 OS=Homo sapiens GN=TNPO3 PE=1 SV=1
248	HNRPK_HUMAN Heterogeneous nuclear ribonucleoprotein K OS=Homo sapiens GN=HNRNPK PE=1 SV=1
244	IMB1_HUMAN Importin subunit beta-1 OS=Homo sapiens GN=KPNB1 PE=1 SV=2
233	H0YN14_HUMAN Importin-4 OS=Homo sapiens GN=IPO4 PE=1 SV=1
233	M0QXS5_HUMAN Heterogeneous nuclear ribonucleoprotein L (Fragment) OS=Homo sapiens GN=HNRNPL PE=1 SV=1
220	SRSF6_HUMAN Serine/arginine-rich splicing factor 6 OS=Homo sapiens GN=SRSF6 PE=1 SV=2
196	I3L3H2_HUMAN Eukaryotic initiation factor 4A-III (Fragment) OS=Homo sapiens GN=EIF4A3 PE=1 SV=1
158	XPOT_HUMAN Exportin-T OS=Homo sapiens GN=XPOT PE=1 SV=2
157	HSP76_HUMAN Heat shock 70 kDa protein 6 OS=Homo sapiens GN=HSPA6 PE=1 SV=2
157	A0A087WVT6_HUMAN Single-stranded DNA-binding protein 3 OS=Homo sapiens GN=SSBP3 PE=1 SV=1
153	BAG2_HUMAN BAG family molecular chaperone regulator 2 OS=Homo sapiens GN=BAG2 PE=1 SV=1
143	A0A075B736_HUMAN Tubulin beta-8 chain OS=Homo sapiens GN=TUBB8 PE=1 SV=1
137	RAGP1_HUMAN Ran GTPase-activating protein 1 OS=Homo sapiens GN=RANGAP1 PE=1 SV=1
126	B1AKQ8_HUMAN Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1 (Fragment) OS=Homo sapiens GN=GNB1 PE=1 SV=5
123	J3KTM9_HUMAN Importin subunit beta-1 (Fragment) OS=Homo sapiens GN=KPNB1 PE=1 SV=1
123	H3BRU6_HUMAN Poly(rC)-binding protein 2 (Fragment) OS=Homo sapiens GN=PCBP2 PE=1 SV=1
122	H0Y3Z3_HUMAN Protein disulfide-isomerase (Fragment) OS=Homo sapiens GN=P4HB PE=1 SV=1
122	ZCCHV_HUMAN Zinc finger CCCH-type antiviral protein 1 OS=Homo sapiens GN=ZC3HAV1 PE=1 SV=3
121	J3KQL8_HUMAN Apolipoprotein L2 OS=Homo sapiens GN=APOL2 PE=1 SV=2
121	H7BY10_HUMAN 60S ribosomal protein L23a (Fragment) OS=Homo sapiens GN=RPL23A PE=1 SV=1
117	PSMD2_HUMAN 26S proteasome non-ATPase regulatory subunit 2 OS=Homo sapiens GN=PSMD2 PE=1 SV=3
113	PFKAL_HUMAN Isoform 2 of ATP-dependent 6-phosphofructokinase, liver type OS=Homo sapiens GN=PFKL
111	LMO4_HUMAN LIM domain transcription factor LMO4 OS=Homo sapiens GN=LMO4 PE=1 SV=1
110	SYAC_HUMAN Alanine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=AARS PE=1 SV=2
108	C9JZR2_HUMAN Catenin delta-1 OS=Homo sapiens GN=CTNND1 PE=1 SV=2
106	PCBP1_HUMAN Poly(rC)-binding protein 1 OS=Homo sapiens GN=PCBP1 PE=1 SV=2

102	MCM7_HUMAN DNA replication licensing factor MCM7 OS=Homo sapiens GN=MCM7 PE=1 SV=4
102	E9PKU4_HUMAN 60S ribosomal protein L8 (Fragment) OS=Homo sapiens GN=RPL8 PE=1 SV=1
101	TRIP6_HUMAN Thyroid receptor-interacting protein 6 OS=Homo sapiens GN=TRIP6 PE=1 SV=3
95	H0YMR4_HUMAN Importin-4 (Fragment) OS=Homo sapiens GN=IPO4 PE=1 SV=1
94	V9GYZ6_HUMAN Bifunctional glutamate/proline--tRNA ligase (Fragment) OS=Homo sapiens GN=EPRS PE=1 SV=1
94	IPO4_HUMAN Importin-4 OS=Homo sapiens GN=IPO4 PE=1 SV=2
91	E9PN89_HUMAN Heat shock cognate 71 kDa protein (Fragment) OS=Homo sapiens GN=HSPA8 PE=1 SV=1
89	DYL2_HUMAN Dynein light chain 2, cytoplasmic OS=Homo sapiens GN=DYNLL2 PE=1 SV=1
86	IF2A_HUMAN Eukaryotic translation initiation factor 2 subunit 1 OS=Homo sapiens GN=EIF2S1 PE=1 SV=3
81	LDB1_HUMAN LIM domain-binding protein 1 OS=Homo sapiens GN=LDB1 PE=1 SV=2
79	I3L397_HUMAN Eukaryotic translation initiation factor 5A (Fragment) OS=Homo sapiens GN=EIF5A PE=1 SV=5
78	NARR_HUMAN Ras-related protein Rab-34, isoform NARR OS=Homo sapiens GN=RAB34 PE=1 SV=1
75	ATD3A_HUMAN Isoform 2 of ATPase family AAA domain-containing protein 3A OS=Homo sapiens GN=ATAD3A
73	HSP7C_HUMAN Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1
73	AT2A2_HUMAN Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 OS=Homo sapiens GN=ATP2A2 PE=1 SV=1
68	E9PKE3_HUMAN Heat shock cognate 71 kDa protein OS=Homo sapiens GN=HSPA8 PE=1 SV=1
66	RBP2_HUMAN E3 SUMO-protein ligase RanBP2 OS=Homo sapiens GN=RANBP2 PE=1 SV=2
65	P2Y12_HUMAN P2Y purinoceptor 12 OS=Homo sapiens GN=P2RY12 PE=1 SV=1
64	H7BZ35_HUMAN Aspartate--tRNA ligase, cytoplasmic (Fragment) OS=Homo sapiens GN=DARS PE=1 SV=1
63	COPA_HUMAN Coatomer subunit alpha OS=Homo sapiens GN=COPA PE=1 SV=2
59	H7C086_HUMAN ATP synthase subunit O, mitochondrial (Fragment) OS=Homo sapiens GN=ATP5O PE=4 SV=1
59	P4HA1_HUMAN Prolyl 4-hydroxylase subunit alpha-1 OS=Homo sapiens GN=P4HA1 PE=1 SV=2
58	E9PLT0_HUMAN Cold shock domain-containing protein E1 OS=Homo sapiens GN=CSDE1 PE=1 SV=1
58	A0A0A0MSG2_HUMAN Four and a half LIM domains protein 2 OS=Homo sapiens GN=FHL2 PE=1 SV=1
58	B2R5W2_HUMAN Heterogeneous nuclear ribonucleoproteins C1/C2 OS=Homo sapiens GN=HNRNPC PE=1 SV=1
57	CSDE1_HUMAN Isoform 3 of Cold shock domain-containing protein E1 OS=Homo sapiens GN=CSDE1
56	A0A0A0MS74_HUMAN Phosphoenolpyruvate carboxykinase [GTP], mitochondrial OS=Homo sapiens GN=PCK2 PE=1 SV=1
55	J3KT29_HUMAN 60S ribosomal protein L23 OS=Homo sapiens GN=RPL23 PE=1 SV=1
52	CAND1_HUMAN Cullin-associated NEDD8-dissociated protein 1 OS=Homo sapiens GN=CAND1 PE=1 SV=2
52	NSD2_HUMAN Histone-lysine N-methyltransferase NSD2 OS=Homo sapiens GN=WHSC1 PE=1 SV=1
52	CCD50_HUMAN Coiled-coil domain-containing protein 50 OS=Homo sapiens GN=CCDC50 PE=1 SV=1
51	Q5QNZ2_HUMAN ATP synthase F(0) complex subunit B1, mitochondrial OS=Homo sapiens GN=ATP5F1 PE=1 SV=1
51	H3BPE7_HUMAN RNA-binding protein FUS OS=Homo sapiens GN=FUS PE=1 SV=1
49	A6NLM8_HUMAN Translocon-associated protein subunit delta OS=Homo sapiens GN=SSR4 PE=1 SV=1
48	ECHA_HUMAN Trifunctional enzyme subunit alpha, mitochondrial OS=Homo sapiens GN=HADHA PE=1 SV=2

48	F8VQQ4_HUMAN Tubulin alpha-1A chain (Fragment) OS=Homo sapiens GN=TUBA1A PE=1 SV=1
47	TBB5_HUMAN Tubulin beta chain OS=Homo sapiens GN=TUBB PE=1 SV=2
45	TBB2A_HUMAN Tubulin beta-2A chain OS=Homo sapiens GN=TUBB2A PE=1 SV=1
45	F6RFD5_HUMAN Destrin OS=Homo sapiens GN=DSTN PE=1 SV=1
44	CSDE1_HUMAN Cold shock domain-containing protein E1 OS=Homo sapiens GN=CSDE1 PE=1 SV=2
43	RS27_HUMAN 40S ribosomal protein S27 OS=Homo sapiens GN=RPS27 PE=1 SV=3
43	MPCP_HUMAN Phosphate carrier protein, mitochondrial OS=Homo sapiens GN=SLC25A3 PE=1 SV=2
43	J3KR97_HUMAN Tubulin-specific chaperone D OS=Homo sapiens GN=TBCD PE=1 SV=1
43	TBB4B_HUMAN Tubulin beta-4B chain OS=Homo sapiens GN=TUBB4B PE=1 SV=1
42	J3KMX5_HUMAN 40S ribosomal protein S13 OS=Homo sapiens GN=RPS13 PE=1 SV=1
42	A0A0D9SFL3_HUMAN RNA-binding protein EWS OS=Homo sapiens GN=EWSR1 PE=4 SV=1
42	H0Y2W2_HUMAN ATPase family AAA domain-containing protein 3A (Fragment) OS=Homo sapiens GN=ATAD3A PE=1 SV=1
40	DNJA1_HUMAN DnaJ homolog subfamily A member 1 OS=Homo sapiens GN=DNAJA1 PE=1 SV=2
40	H0YDD4_HUMAN Acetyltransferase component of pyruvate dehydrogenase complex (Fragment) OS=Homo sapiens GN=DLAT PE=1 SV=1
40	STRAP_HUMAN Serine-threonine kinase receptor-associated protein OS=Homo sapiens GN=STRAP PE=1 SV=1
39	CY1_HUMAN Cytochrome c1, heme protein, mitochondrial OS=Homo sapiens GN=CYC1 PE=1 SV=3
39	H0Y8N0_HUMAN Transmembrane protein 33 (Fragment) OS=Homo sapiens GN=TMEM33 PE=1 SV=1
39	TBA1B_HUMAN Tubulin alpha-1B chain OS=Homo sapiens GN=TUBA1B PE=1 SV=1
38	IF4A3_HUMAN Eukaryotic initiation factor 4A-III OS=Homo sapiens GN=EIF4A3 PE=1 SV=4
38	Q5JR08_HUMAN Rho-related GTP-binding protein RhoC (Fragment) OS=Homo sapiens GN=RHOC PE=1 SV=5
38	F5H5D3_HUMAN Tubulin alpha-1C chain OS=Homo sapiens GN=TUBA1C PE=1 SV=1
38	P4HA2_HUMAN Prolyl 4-hydroxylase subunit alpha-2 OS=Homo sapiens GN=P4HA2 PE=1 SV=1
37	SND1_HUMAN Staphylococcal nuclease domain-containing protein 1 OS=Homo sapiens GN=SND1 PE=1 SV=1
36	TBA1A_HUMAN Tubulin alpha-1A chain OS=Homo sapiens GN=TUBA1A PE=1 SV=1
36	I3L404_HUMAN 40S ribosomal protein S2 (Fragment) OS=Homo sapiens GN=RPS2 PE=1 SV=1
34	V9GYM8_HUMAN Rho guanine nucleotide exchange factor 2 OS=Homo sapiens GN=ARHGEF2 PE=1 SV=1
34	RAP1B_HUMAN Isoform 4 of Ras-related protein Rap-1b OS=Homo sapiens GN=RAP1B
34	ATD3B_HUMAN ATPase family AAA domain-containing protein 3B OS=Homo sapiens GN=ATAD3B PE=1 SV=1
33	TBA4A_HUMAN Tubulin alpha-4A chain OS=Homo sapiens GN=TUBA4A PE=1 SV=1
33	E9PI65_HUMAN Heat shock cognate 71 kDa protein (Fragment) OS=Homo sapiens GN=HSPA8 PE=1 SV=1
32	CNN3_HUMAN Calponin-3 OS=Homo sapiens GN=CNN3 PE=1 SV=1
30	J3KTJ3_HUMAN 60S ribosomal protein L23 OS=Homo sapiens GN=RPL23 PE=1 SV=1
30	TBA3C_HUMAN Tubulin alpha-3C/D chain OS=Homo sapiens GN=TUBA3C PE=1 SV=3
30	PCKGM_HUMAN Phosphoenolpyruvate carboxykinase [GTP], mitochondrial OS=Homo sapiens GN=PCK2 PE=1 SV=3
29	AP2A2_HUMAN AP-2 complex subunit alpha-2 OS=Homo sapiens GN=AP2A2 PE=1 SV=2
29	Q5T4U5_HUMAN Acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain, isoform CRA_a OS=Homo sapiens GN=ACADM PE=1 SV=1
27	A0A0A0MSX9_HUMAN Isoleucine--tRNA ligase, cytoplasmic OS=Homo sapiens GN=IARS PE=1 SV=1
27	E5RIM3_HUMAN Phospholipase A-2-activating protein OS=Homo sapiens GN=PLAA PE=1 SV=1

27	C9J2C0_HUMAN Tubulin alpha-8 chain (Fragment) OS=Homo sapiens GN=TUBA8 PE=1 SV=1
27	A2A3R7_HUMAN 40S ribosomal protein S6 OS=Homo sapiens GN=RPS6 PE=1 SV=1
26	F5H2T0_HUMAN Elongator complex protein 1 OS=Homo sapiens GN=IKBKAP PE=1 SV=1
26	LPPRC_HUMAN Leucine-rich PPR motif-containing protein, mitochondrial OS=Homo sapiens GN=LRPPRC PE=1 SV=3
25	X1WI28_HUMAN 60S ribosomal protein L10 (Fragment) OS=Homo sapiens GN=RPL10 PE=1 SV=4
24	RPN1_HUMAN Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 OS=Homo sapiens GN=RPN1 PE=1 SV=1
23	KHDR1_HUMAN KH domain-containing, RNA-binding, signal transduction-associated protein 1 OS=Homo sapiens GN=KHDRBS1 PE=1 SV=1
23	J3KQE5_HUMAN GTP-binding nuclear protein Ran (Fragment) OS=Homo sapiens GN=RAN PE=1 SV=1
22	B4DR61_HUMAN Protein transport protein Sec61 subunit alpha isoform 1 OS=Homo sapiens GN=SEC61A1 PE=1 SV=1
21	SPG20_HUMAN Spartin OS=Homo sapiens GN=SPG20 PE=1 SV=1
21	GSH0_HUMAN Glutamate--cysteine ligase regulatory subunit OS=Homo sapiens GN=GCLM PE=1 SV=1
21	HS90B_HUMAN Heat shock protein HSP 90-beta OS=Homo sapiens GN=HSP90AB1 PE=1 SV=4
21	EFTU_HUMAN Elongation factor Tu, mitochondrial OS=Homo sapiens GN=TUFM PE=1 SV=2
20	M0QXL5_HUMAN rRNA 2'-O-methyltransferase fibrillarin (Fragment) OS=Homo sapiens GN=FBL PE=1 SV=1
20	K7EK77_HUMAN ATP synthase subunit alpha, mitochondrial (Fragment) OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
20	HUWE1_HUMAN E3 ubiquitin-protein ligase HUWE1 OS=Homo sapiens GN=HUWE1 PE=1 SV=3
20	ADT2_HUMAN ADP/ATP translocase 2 OS=Homo sapiens GN=SLC25A5 PE=1 SV=7
19	ATPA_HUMAN ATP synthase subunit alpha, mitochondrial OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
19	RS14_HUMAN 40S ribosomal protein S14 OS=Homo sapiens GN=RPS14 PE=1 SV=3
19	GRP75_HUMAN Stress-70 protein, mitochondrial OS=Homo sapiens GN=HSPA9 PE=1 SV=2
19	GNAL_HUMAN Guanine nucleotide-binding protein G(olf) subunit alpha OS=Homo sapiens GN=GNAL PE=1 SV=1
19	A0A0D9SEM4_HUMAN Serine/arginine-rich-splicing factor 4 (Fragment) OS=Homo sapiens GN=SRSF4 PE=4 SV=1
19	RINI_HUMAN Ribonuclease inhibitor OS=Homo sapiens GN=RNH1 PE=1 SV=2
19	HS90A_HUMAN Heat shock protein HSP 90-alpha OS=Homo sapiens GN=HSP90AA1 PE=1 SV=5
18	ATPA_HUMAN Isoform 3 of ATP synthase subunit alpha, mitochondrial OS=Homo sapiens GN=ATP5A1
18	H2AZ_HUMAN Histone H2A.Z OS=Homo sapiens GN=H2AFZ PE=1 SV=2
17	F8W026_HUMAN Endoplasmin (Fragment) OS=Homo sapiens GN=HSP90B1 PE=1 SV=5
17	K7ENJ4_HUMAN ATP synthase subunit alpha, mitochondrial (Fragment) OS=Homo sapiens GN=ATP5A1 PE=1 SV=1
17	H0YBW4_HUMAN Phospholipase A-2-activating protein (Fragment) OS=Homo sapiens GN=PLAA PE=1 SV=1
16	B4DXZ6_HUMAN Fragile X mental retardation syndrome-related protein 1 OS=Homo sapiens GN=FXR1 PE=1 SV=1
16	SRRM2_HUMAN Serine/arginine repetitive matrix protein 2 OS=Homo sapiens GN=SRRM2 PE=1 SV=2
16	I3L0K7_HUMAN Heat shock protein 75 kDa, mitochondrial OS=Homo sapiens GN=TRAP1 PE=1 SV=1
16	CH60_HUMAN 60 kDa heat shock protein, mitochondrial OS=Homo sapiens GN=HSPD1 PE=1 SV=2
16	E9PEJ4_HUMAN Acetyltransferase component of pyruvate dehydrogenase complex OS=Homo sapiens GN=DLAT PE=1 SV=1
16	B0YJC4_HUMAN Vimentin OS=Homo sapiens GN=VIM PE=1 SV=1
15	GRP78_HUMAN 78 kDa glucose-regulated protein OS=Homo sapiens GN=HSPA5 PE=1 SV=2

15	EF2_HUMAN Elongation factor 2 OS=Homo sapiens GN=EEF2 PE=1 SV=4
15	COPB_HUMAN Coatomer subunit beta OS=Homo sapiens GN=COPB1 PE=1 SV=3
14	A0A087WVQ9_HUMAN Elongation factor 1-alpha 1 OS=Homo sapiens GN=EEF1A1 PE=1 SV=1
14	TIF1B_HUMAN Transcription intermediary factor 1-beta OS=Homo sapiens GN=TRIM28 PE=1 SV=5
14	G3V1A4_HUMAN Cofilin 1 (Non-muscle), isoform CRA_a OS=Homo sapiens GN=CFL1 PE=1 SV=1
14	G8JLB6_HUMAN Heterogeneous nuclear ribonucleoprotein H OS=Homo sapiens GN=HNRNPH1 PE=1 SV=1
14	E7EMW7_HUMAN E3 ubiquitin-protein ligase UBR5 OS=Homo sapiens GN=UBR5 PE=1 SV=1
14	CO4A_HUMAN Complement C4-A OS=Homo sapiens GN=C4A PE=1 SV=2
14	FAS_HUMAN Fatty acid synthase OS=Homo sapiens GN=FASN PE=1 SV=3
14	J3JS69_HUMAN 40S ribosomal protein S18 OS=Homo sapiens GN=RPS18 PE=1 SV=1
13	RS18_HUMAN 40S ribosomal protein S18 OS=Homo sapiens GN=RPS18 PE=1 SV=3
13	H3BRG4_HUMAN Cytochrome b-c1 complex subunit 2, mitochondrial OS=Homo sapiens GN=UQCRC2 PE=1 SV=1
13	H0Y368_HUMAN Dolichol-phosphate mannosyltransferase subunit 1 (Fragment) OS=Homo sapiens GN=DPM1 PE=1 SV=1
13	POTEE_HUMAN POTE ankyrin domain family member E OS=Homo sapiens GN=POTEE PE=1 SV=3
13	F5H018_HUMAN GTP-binding nuclear protein Ran (Fragment) OS=Homo sapiens GN=RAN PE=1 SV=5
13	ACTC_HUMAN Actin, alpha cardiac muscle 1 OS=Homo sapiens GN=ACTC1 PE=1 SV=1
12	DYHC1_HUMAN Cytoplasmic dynein 1 heavy chain 1 OS=Homo sapiens GN=DYNC1H1 PE=1 SV=5
12	B4DEB1_HUMAN Histone H3 OS=Homo sapiens GN=H3F3A PE=1 SV=1
12	LR16A_HUMAN Leucine-rich repeat-containing protein 16A OS=Homo sapiens GN=LRRC16A PE=1 SV=1
12	RL27_HUMAN 60S ribosomal protein L27 OS=Homo sapiens GN=RPL27 PE=1 SV=2
12	ATPG_HUMAN ATP synthase subunit gamma, mitochondrial OS=Homo sapiens GN=ATP5C1 PE=1 SV=1
12	WDR1_HUMAN WD repeat-containing protein 1 OS=Homo sapiens GN=WDR1 PE=1 SV=4
12	ATPB_HUMAN ATP synthase subunit beta, mitochondrial OS=Homo sapiens GN=ATP5B PE=1 SV=3
12	EF1A3_HUMAN Putative elongation factor 1-alpha-like 3 OS=Homo sapiens GN=EEF1A1P5 PE=5 SV=1
12	A0A087WV01_HUMAN Elongation factor 1-alpha OS=Homo sapiens GN=EEF1A1 PE=1 SV=1
11	A0A087X0X3_HUMAN Heterogeneous nuclear ribonucleoprotein M OS=Homo sapiens GN=HNRNPM PE=1 SV=1
11	E7EQV9_HUMAN Ribosomal protein L15 (Fragment) OS=Homo sapiens GN=RPL15 PE=1 SV=1
11	A0A087WZ27_HUMAN Zinc finger protein 90 OS=Homo sapiens GN=ZNF90 PE=4 SV=2
11	SRSF7_HUMAN Isoform 2 of Serine/arginine-rich splicing factor 7 OS=Homo sapiens GN=SRSF7
10	SERA_HUMAN D-3-phosphoglycerate dehydrogenase OS=Homo sapiens GN=PHGDH PE=1 SV=4

N.B. Ratios are mean(Unr)/mean(IgG) and all infinities were ignored unless every Unr value was greater than zero. The yellow shading highlights a protein absent in every IgG sample but present in every Unr sample. Finite ratios were rounded to the nearest whole number.